PB91-154591

Anticipated Pesticide Residues in Food. Volume 1

Dynamac Corp., Rockville, MD

Prepared for:

Environmental Protection Agency, Washington, DC

26 Jul 89

U.S. Department of Commerce
National Technical Information Service
NTIS

PB91-154591

ANTICIPATED PESTICIDE RESIDUES IN FOOD

EPA Contract No. 68-01-7363

Work Assignment No. 54

DRAFT FINAL REPORT

Volume 1

July 26, 1989

SUBMITTED TO:

Joseph Reinert
Office of Policy, Planning and Evaluation
U.S. Environmental Protection Agency
401 M Street, SW
Washington, DC 20460

SUBMITTED BY:

Dynamac Corporation
The Dynamac Building
11140 Rockville Pike
Rockville, MD 20852

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA 22161

BIBLIOGRAPHIC INFORMATION

PB91-154591

Report Nos: none

Title: Anticipated Pesticide Residues in Food. Volume 1.

Date: 26 Jul 89

Authors: J. Reinert.

Performing Organization: Dynamac Corp., Rockville, MD.

Sponsoring Organization: *Environmental Protection Agency, Washington, DC. Office of Policy, Planning and Evaluation.

Contract Nos: EPA-68-Ø1-7363

Type of Report and Period Covered: Draft rept. (Final).

Supplementary Notes: See also Volume 2, PB91-154609. Sponsored by Environmental Protection Agency, Washington, DC. Office of Policy, Planning and Evaluation.

NTIS Field/Group Codes: 98H, 68E

Price: PC AØ5/MF AØ1

Availability: Available from the National Technical Information Service, Springfield, VA. 22161

Number of Pages: 99p

Keywords: *Pesticide residues, *Food contamination, Monitoring, Data bases, Tables(Data), United States, Canada, Food processing, Data acquisition, Data processing, Access methods.

Abstract: Pesticide residue data on raw and processed food were obtained on a voluntary basis from agencies and organizations not required to submit these data to the Office of Pesticide Programs (OPP). The project was part of a scoping exercise to compile and summarize in an electronic format available monitoring data from outside the Agency, and to begin to evaluate the usefulness of these data for scientific and regulatory purposes. Data for 286 pesticides on an estimated 49,857 samples were obtained from three sources: Agriculture Canada, state monitoring data compiled by FDA, and the National Food Processors Association.

PB91-154591

ANTICIPATED PESTICIDE RESIDUES IN FOOD

EPA Contract No. 68-01-7363

Work Assignment No. 54

DRAFT FINAL REPORT

Volume 1

July 26, 1989

SUBMITTED TO:

Joseph Reinert
Office of Policy, Planning and Evaluation
U.S. Environmental Protection Agency
401 M Street, SW
Washington, DC 20460

SUBMITTED BY:

Dynamac Corporation
The Dynamac Building
11140 Rockville Pike
Rockville, MD 20852

TABLE OF CONTENTS

VOLUME 1

	Acknowledgments	ii
	Executive Summary	iii
A.	Introduction	1
B.	Data Sources	2
C.	Project Data Base Construction	10
D.	Data Evaluation	20
E.	Proposed Standard Data Coding System and Transmission Format	54
F.	Conclusions and Recommendations	59

Appendix A: Example Search of the Project Data Base

Appendix B: Letter of Questions to FDA

Appendix C: Oracle RDBMS - Technical Evaluation

Appendix D: Project Data Base Design

VOLUME 2

Data Listings from Project Data Base

1. Agriculture Canada Data
2. National Food Processors Association Data
3. FoodContam Pesticides Look-Up Table
4. FoodContam Food Products Look-Up Table

VOLUME 3

Data Listings from Project Data Base

FoodContam Data : Records 1 - 20,000

VOLUME 4

Data Listings from Project Data Base

FoodContam Data : Records 20,001 - 42,974

ACKNOWLEDGMENTS

This report was prepared by a multidisciplinary team of Dynamac scientists and information specialists. Dr. J. Dirk Nies was Task Manager. Other members of the team were:

Peter Clamp, principal author/Oracle RDBMS specialist;
Brion Cook, biostatistician/data format specialist; and
Rachel Spangenberg, author/researcher.

The authors especially wish to thank Bob Dickinson, FDA, and Jim Minyard, Mississippi State, for supplying the 1988 FoodContam State pesticide residue data; Ed Elkins and John Heuser, NFPA, for supplying industry-generated data; and Gil Flores of Agriculture Canada for providing the Canadian data.

EXECUTIVE SUMMARY

Pesticide residue data on raw and processed food were obtained on a voluntary basis from agencies and organizations not required to submit these data to the Office of Pesticide Programs (OPP). ~~This~~ The project was part of a scoping exercise to compile and summarize in an electronic format available monitoring data from outside the Agency, and to begin to evaluate the usefulness of these data for scientific and regulatory purposes.

Data for 286 pesticides on an estimated 49,857 samples were obtained from three sources: Agriculture Canada, state monitoring data compiled by FDA, and the National Food Processors Association. The number of distinct samples and the number of distinct pesticides represented in each data set were estimated as follows:

DATA SET	SAMPLES	PESTICIDES
Agriculture Canada	5,096	153
FDA state monitoring data	17,012	155
National Food Processors Association	27,749	135

The data were edited to remove obvious errors and inconsistencies, and were then entered into tables using the Oracle Relational Data Base Management System (RDBMS). These tables (with the addition of 'indexes' and 'views' that make the data more accessible) form the 'Project Data Base', which was analyzed to make further tests of consistency and to generate summary statistics. This report records these tests and analyses, and discusses the uses and limitations of the data for estimating pesticide residues in the general food supply. Recommendations are made for adoption of the following:

1. Standard residue sampling protocols used by state and federal agencies to improve the comparability and usefulness of data;

2. A minimum set of information to be provided with each sample (e.g., sample ID, purpose of sample, analytical method); and

3. A standard data coding system and data base format for the transmission of data among laboratories and the agencies and organizations responsible for monitoring the safety of our food supply.

In our judgment, the adoption of these recommendations would substantially improve the utility of the pesticide residue data currently being generated without significantly increasing the cost of obtaining and reporting these data.

A. INTRODUCTION

The OPP's dietary assessments for pesticides are consistently hampered by the lack of good information on pesticide residues in foods at the time of consumption. These anticipated residues (ARs) should ideally consider the effects of storage, processing, and cooking on residues consumed.

Tolerances are enforcement tools and are ill-suited for risk assessment. Theoretically, the field trial data submitted to support tolerances could be used for dietary assessments. This would require estimating the percent of crop treated and other use information to extrapolate to the actual dietary burden. Such estimates are often difficult to derive. Also, data on the effects of food storage, processing, and cooking on residue levels are typically unavailable in OPP, making estimations of ARs very difficult. AR data reliability is even more critical when dietary exposure information is used for chronic health assessments such as oncogenicity.

Under the previous Work Assignment No. 19, the Office of Policy, Planning and Evaluation undertook a broad-scale effort to obtain and summarize monitoring information available on pesticide residues present in raw agricultural commodities, animal products, and processed foodstuffs developed by other governmental agencies and food processing companies. Data routinely available to OPP, e.g., FDA monitoring data, were not included in this project because it would have been resource intensive and redundant to do so. The analytical accuracy and precision of the data were assessed. The structure of a Lotus 1-2-3 spreadsheet was established to store this information, and much of the data entry was completed. This project was initiated to obtain additional residue information, to complete construction of a composite data base, and to perform analyses of these data.

B. DATA SOURCES

Data were obtained in an electronic format from three sources:

1. the Food Residue Monitoring Consolidation Program (FRMCP) coordinated by the Pesticides Directorate, Agriculture Canada;

2. the Food and Drug Administration's FoodContam Data Base and Data Entry System, provided by the Mississippi State Chemical Laboratory; and

3. the National Food Processors Association (NFPA).

Figures 1, 2, and 3 in this section give examples of data records essentially in the form in which they were received. Volumes 2, 3, and 4 include the full data sets from all three sources after editing by Dynamac. The data analyses presented in section D were performed directly from the Oracle RDBMS tables that form the Project Data Base. An example of how to search the Project Data Base for a specific pesticide residue is provided in Appendix A. An example of questions raised to the data submitters addressing the utility and reliability of their data appears in Appendix B.

B.1. Data Overview

B.1.1 Agriculture Canada

Data from Agriculture Canada were provided as a single Wang List Processing file containing 2,825 records. Each record gives data for a batch containing from 1 to 501 samples of the same food product for which analytical results are reported for a single pesticide. A total of 153 different pesticides are referenced. Samples were collected from Canadian products between April 1986 and March 1987 (Canadian Grain Commission from August 1985 to July 1986). No imported commodities were included.

Data fields in the original file are as follows:

Pesticide Analyzed
Food Commodity
Agency
% in Compliance with MRL
Maximum Residue Limit (MRL) (ppm)

Lowest Reporting Level (ppm)
Total Samples
No. of Samples with No Detected Residues
Range of Detected Levels (ppm)
Average Detected (ppm)

Some typical Agriculture Canada data records are shown in Figure 1.

B.1.2 FoodContam

Data from FoodContam were provided on three 1.2-megabyte floppy diskettes as archived files. These files were un-archived using the utility program provided by Mississippi State to yield a total of 42,974 records in ASCII format. Each record includes a sample ID field. However, many samples are reported in more than one record; only 17,012 distinct sample IDs were counted. A total of 155 different pesticides are recorded. Samples were collected in FY 1988 from seven states: California, Florida, Massachusetts, Michigan, New York, Virginia, and Wisconsin.

The following data fields are used by FoodContam:

Analytical Lab Reporting
Date Report Submitted
Product Code
Collect Date
Report Date
State
County
Sample Number
Detection Method Code
Extraction Method Code
Analyte Code
Amount Found
Significant
Product Common Name

Some typical FoodContam data records are shown in Figure 2. Records are stored in an efficient format that includes fixed-length fields with no spaces. Four fields contain codes: pesticide/analyte, food product, determination method, and extraction method are each coded using predefined lists of codes.

Figure 1. Typical Agriculture Canada Data Records

Pesticide Analysed	Food Commodity	Agency	% in Compl. with MRL	Maximum Residue Limit (MRL) (ppm)	Lowest Reporting Level (ppm)	Total Samples	No. of Samples with No Detected Residues	Range of Detected Levels (ppm)	Average Detected (ppm)
2,4,5-T(4)	Feed	Feed/Fert	. .	. .	0.01	35	35		
2,4-D(4)	Feed	Feed/Fert			0.01	35	35		
2,4-DP(10)	Cereal Grains(10)	Can Grain Comm	100	0.10*	0.010	227	227		
2,4-DP(10)	RapeseedCanola(12)	Can Grain Comm	**	0.10*	0.02	127	127		
2,4-DP(10)	Soybeans	Can Grain Comm	100	0.10	0.010	7	7		
Acephate	Apricots	BC Agr/Fish	100	x	0.1	10	10		
Acephate	Beef	Health Protection	100	0.1	0.1	58	58	x	x
Acephate	Blueberries	BC Agr/Fish	100	x	0.1	11	11		
Acephate	Cabbage	BC Agr/Fish	100	x	0.1	6	6		
Acephate	Cauliflower	BC Agr/Fish	100	x	0.1	15	15		
Acephate	Celery	BC Agr/Fish	100	x	0.1	13	13		
Acephate	Cherries	BC Agr/Fish	100	x	0.1	7	7		
Acephate	Chinese Veg	BC Agr/Fish	100	x	0.1	15	15		
Acephate	Cucumber, Greenhouse	BC Agr/Fish	100	x	0.1	5	5		
Acephate	Grapes	BC Agr/Fish	100	x	0.1	20	20		
Acephate	Lettuce	BC Agr/Fish	100	x	0.1	20	20		
Acephate	Mushrooms	BC Agr/Fish	100	x	0.1	20	20		
Acephate	Onion, Green	BC Agr/Fish	100	x	0.1	10	10		
Acephate	Pork	Health Protection	100	0.1	0.1	34	34	x	x
Acephate	Spinach	BC Agr/Fish	100	x	0.1	5	5		
Acephate	Tomatoes, Field	BC Agr/Fish	100	x	0.1	10	10		
Acephate	Tomatoes, Greenhouse	BC Agr/Fish	100	x	0.1	5	5		
Acephate	Turnip	BC Agr/Fish	100	x	0.1	--	--		
Alachlor	Apricots	BC Agr/Fish	100	x	0.5	10	10		
Alachlor	Blueberries	BC Agr/Fish	100	x	0.05	11	11		
Alachlor	Cabbage	BC Agr/Fish	100	x	0.05	6	6		
Alachlor	Cauliflower	BC Agr/Fish	100	x	0.5	15	15		
Alachlor	Celery	BC Agr/Fish	100	x	0.5	13	13		
Alachlor	Cherries	BC Agr/Fish	100	x	0.5	7	7		
Alachlor	Chinese Veg	BC Agr/Fish	100	x	0.5	15	15		
Alachlor	Corn	Health Protection	100	0.1	0.04	99	99	x	x
Alachlor	Cucumber, Greenhouse	BC Agr/Fish	100	x	0.5	5	5		
Alachlor	Grapes	BC Agr/Fish	100	x	0.05	20	20		
Alachlor	Lettuce	BC Agr/Fish	100	x	0.5	20	20		
Alachlor	Mushrooms	BC Agr/Fish	100	x	0.5	10	10		
Alachlor	Onion, Green	BC Agr/Fish	100	x	0.5	10	10		
Alachlor	Spinach	BC Agr/Fish	100	x	0.5	5	5		
Alachlor	Tomatoes, Field	BC Agr/Fish	100	x	0.5	10	10		
Alachlor	Tomatoes, Greenhouse	BC Agr/Fish	100	x	0.5	5	5		
Alachlor	Turnip	BC Agr/Fish	100	x	0.5	--	--		
Aldicarb	Apricots	BC Agr/Fish	100	x	0.02	11	11		
Aldicarb	Blueberries	BC Agr/Fish	100	x	0.02	6	6		
Aldicarb	Cabbage	BC Agr/Fish	100	x	0.02	15	15		
Aldicarb	Cauliflower	BC Agr/Fish	100	x	0.2	13	13		
Aldicarb	Celery	BC Agr/Fish	100	x	0.2				

Figure 2. Typical FoodContam Data Records

```
0415A01870904871001CA0592000718901999000      EGGS
0415A01870904871001CA0592000718901999160.200  EGGS
0415A01870904871001CA0592000718916999024.300  EGGS
0415A01870904871001CA0592000718916999170.030  EGGS
0415A01870904871001CA0592000718959999000      EGGS
0415A01871001871001CA0712007190059999255O.00  EGGS
0415A01871001871001CA0732007191059999255O.00  EGGS
0415A01871020871020CA1112007226059999255O.00  EGGS
0420A01870928870929CA0372000621901999000      RED RASPBERRY
0420A01870928870929CA0372000621916999000      RED RASPBERRY
0420A01870928870929CA0372000621959999000      RED RASPBERRY
0420A02870708870710CA0231007193005999000      BLUEBERRY
0420A02870708870710CA0231007193016999000      BLUEBERRY
0420A02870724870724CA0372000547801999000      BLUEBERRY
0420A02870724870724CA0372000547816999000      BLUEBERRY
0420A02870724870724CA0372000547859999000      BLUEBERRY
0420A02870806870807CA0131007223602999000      BLUEBERRY
0420A02870806870807CA0131007223605999000      BLUEBERRY
0420A02870806870807CA0131007223616999000      BLUEBERRY
0420A02870826870826CA0372000587801999000      BLUEBERRY
0420A02870826870826CA0372000587816999000      BLUEBERRY
0420A02870826870826CA0372000587859999000      BLUEBERRY
0420A04871020871020CA0372000655801999164.280  CRANBERRY
0420A04871020871020CA0372000655816999000O.00  CRANBERRY
0420A04871020871020CA0372000655859999000O.00  CRANBERRY
0420A04871116871117CA0674003131501999164.040  CRANBERRY
0420A04871116871117CA0674003131502999000O.00  CRANBERRY
0420A04871116871117CA0674003131512999000O.00  CRANBERR.
0420A04871223871224CA0394003160901999000O.00  CRANBERRY
0420A04871223871224CA0394003160902999000O.00  CRANBERRY
0420A04871223871224CA0394003160912999000O.00  CRANBERRY
0420A09870708870709CA0654002936101999000      GRAPES(ALL OR UNSPEC)
0420A09870708870709CA0654002936102999000      GRAPES(ALL OR UNSPEC)
0420A09870708870709CA0654002936159999000      GRAPES(ALL OR UNSPEC)
0420A09870709870710CA0011007193505999000      GRAPES(ALL OR UNSPEC)
0420A09870709870710CA0011007193516999000      GRAPES(ALL OR UNSPEC)
0420A09870713870714CA0011007195202999000      GRAPES(ALL OR UNSPEC)
0420A09870713870714CA0011007195205999000      GRAPES(ALL OR UNSPEC)
0420A09870713870714CA0011007195216999000      GRAPES(ALL OR UNSPEC)
0420A09870713870714CA0372000531901999000      GRAPES(ALL OR UNSPEC)
0420A09870713870714CA0372000531916999000      GRAPES(ALL OR UNSPEC)
0420A09870713870714CA0372000531959999000      GRAPES(ALL OR UNSPEC)
0420A09870715870715CA0851007196702999000      GRAPES(ALL OR UNSPEC)
0420A09870715870715CA0851007196705999000      GRAPES(ALL OR UNSPEC)
0420A09870715870715CA0851007196716999000      GRAPES(ALL OR UNSPEC)
0420A09870723870724CA0372000547501999000      GRAPES(ALL OR UNSPEC)
0420A09870723870724CA0372000547516999000      GRAPES(ALL OR UNSPEC)
0420A09870723870724CA0372000547559999000      GRAPES(ALL OR UNSPEC)
0420A09870724870724CA0372000547401999000      GRAPES(ALL OR UNSPEC)
```

In addition to the data itself, Mississippi State supplied to Dynamac a copy of an automated Data Entry System on a single PC-compatible floppy disc. The system includes an executable program and a large number of look-up files. The program presents the user with a data entry screen and performs some interactive data verification. Entries in the four coded fields may be made directly or by means of menus (constructed from the look-up files), which include both the codes and the full names of pesticides, food products, detection methods, and extraction methods.

For this study, the look-up files have been adopted as part of the FoodContam data set. Although not supplied for this purpose, the look-up files have been merged (where necessary), edited, and read into Oracle look-up tables. In the context of the FoodContam Data Entry System, the look-up files serve the function of compressing text descriptions into standard codes; in the Project Data Base, the Oracle look-up tables serve the reverse function of allowing the codes to be translated back into their textual descriptions.

B.1.3 National Food Processors Association

Data from NFPA were provided on a single floppy diskette in six Lotus 1-2-3 files containing a total of 1,865 records. Each record gives data for a batch containing from 1 to 1,609 samples of the same food product for which analytical results are reported for a single pesticide. The data include 136 different pesticides and were collected from 1980 to 1988. Data were supplied by 17 member companies of NFPA.

Data fields in the original Lotus files are headed as follows:

Company
Lab
Form
Processed
Baby Food
Tolerance
Pesticide
Year
Method
#Samples
#Samples<DL
Mean

D.L. Ref
Min. Range
Max. Range
Region
Source
PestApplHist
Recovery Ref
Amt. Prod

Some typical NFPA data records are shown in Figure 3; the fields 'Company' to '#Samples<DL' are shown on the first page, and the fields 'Mean' to 'Amt. Prod' are shown on the second page.

Figure 3. Typical NFPA Data Records (page 1 of 2)

Company	Lab	Form	Processed	Baby Food	Tolerance	Pesticide	Year	Method	#Samples	#Samples<DL	
	16	Bananas	p	Yes	0.3	Aldicarb	1987		81		71
	16	Bananas	p	Yes	0.3	Aldicarb	1988		17		17
	16	Canned Bananas	p	Yes	0.3	Aldicarb	1987		16		13
	7	Orange Conc.	p	n	.	Benomyl	1986		11		11
	11	Peaches	p	n	15	benomyl			6		3
	14	Grape Juice	p	n	10	benomyl	1987		12		11
	2	Apple Paste	p	Yes	7	Benomyl	1987		1		0
	16	Canned Bananas	p	Yes	0.2	Benomyl	1987		3		3
	2	Canned Peaches	p	Yes	15	Benomyl	1987		1		0
	16	Apricot Juice Concp		Yes	20	Botran	1984		3		3
	16	Apricot Juice Concp		Yes	20	Botran	1095	NRA	8		7
	16	Apricot Juice Concp		Yes	20	Botran	1986	NRA	5		50
	16	Apricot Juice Concp		Yes	20	Botran	1987	NRA	64		41
	16	Apricot Juice Concp		Yes	20	Botran	1983	NRA	13		13
	16	Peach Conc.	p	Yes	20	Botran	1986	NRA	36		36
	16	Peach Conc.	p	Yes	20	Botran	1987	NRA	28		28
	8	Orange Juice	p	n	.	Captafol	1987	NRA	4		4
	8	Grapefruit	p	n	.	Captafol	1987	NRA	2		2
	12	Orange Conc.	p	n	.	captafol	1986				
	7	Orange Conc.	p	n	25	Captan	1986	NRA	11		11
	8	Orange Juice	p	n	25	Captan	1987	NRA	4		4
	8	Grapefruit	p	n	25	Captan	1987	NRA	2		2
	14	Grape Juice	p	n	50	Captan	1985		21		8
	14	Grape Juice	p	n	50	Captan	1985		8		8
	10	Apple Juice	p	YES	25	Captan	1984	NRA	178		177
	16	Cooked Apples	p	Yes	25	Captan	1987	NRA			
	10	Juice Concentratesp		Yes		Captan	1987	NRA	250		229
	10	Juice Concentratesp		Yes		Captan	1986	NRA	307		302
	10	Juice Concentratesp		Yes		Captan	1985	NRA	236		231
	10	Mixed Fruit	p	YES		Captan	1984	NRA	108		107
	16	Peaches Cooked	p	Yes	50	Captan	1987	NRA	3		3
	16	Pears Finished	p	Yes	25	Captan	1987	NRA	2		2
	10	Plums Frozen	p	Yes	50	Captan	1986	NRA	20		9
	10	Plums Frozen	p	Yes	50	Captan	1987	NRA	17		7
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1980	NRA	18		17
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1981	NRA	23		23
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1982	NRA	27		27
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1983	NRA	30		30
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1984	NRA	22		22
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1985	NRA	19		19
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1986	NRA	18		18
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1987	NRA	71		71
	16	Cherry Juice Conc.	p	Yes	0.1	Chlorodane	1988	NRA	41		41
	8	Orange Juice	p	n	.	Chlorothalonil	1987	NRA	4		2
	8	Grapefruit	p	n	.	Chlorothalonil	1987	NRA	2		1
	10	Juice Concentratesp		Yes		Cygon	1985	NRA	236		233
	16	Apple Juice Conc.	p	Yes	0.1	DDE	1980	NRA	2		2
	16	Apple Juice Conc.	p	Yes	0.1	DDE	1981	NRA	3		3
	16	Apple Juice Conc.	p	Yes	0.1	DDE	1982	NRA	11		11
	16	Apple Juice Conc.	p	Yes	0.1	DDE	1983	NRA	11		11

Figure 3. Typical NFPA Data Records (page 2 of 2)

Main	D.L.Ref	Min.Range	Max.Range	Region	Source	PestApplHist	Recovery Ref	Ant.Prod
0.07		0.05	0.005	0.019				
0.06		0.01						
0.101		0.01						
0.01		0.01						
0.39		0.01						
0.01		0.1						
		0.1						
0.05		0.1						
		0.1						
0.1		0.1						
0.1		0.1						
		0.05	0.05	0.43				
		0.01						
		0.03						
		0.03						
	Tr.		0.05	0.12				
	Tr.			0.04				
				0.03				
			0.008	0.04				
	Tr.			0.3				
	Tr.			0.038				
		0.05						
		0.05						
		0.05	

C. PROJECT DATA BASE CONSTRUCTION

C.1 Oracle Relational Data Base Management System

During construction of the pesticide residue data base, it became apparent that Lotus 1-2-3 (the format in which the NFPA data were submitted) was not capable of storing and manipulating the quantity of data obtained from other sources, and that a more powerful data base management system was required. The Oracle Relational Data Base Management System (RDBMS) was chosen because it has superior data storage and retrieval capabilities, and because it is compatible across more than 30 operating systems and 80 hardware platforms, including the IBM PC-AT and Apple Macintosh.

Under Oracle, information is stored in tables of rows (records) and columns (fields). Relationships may be established between different tables that include 'link' columns containing similar information. For example, in this study a 'look-up' table of pesticide codes and names has been used to add pesticide names to a report based on coded records of tests. A further table could be linked that lists tolerances for given pesticide/use combinations; other similar tables may be required to meet specific needs for data analysis. The advantage of relating data in separate tables is that information is stored economically and naturally and can be retrieved and edited easily.

Oracle data base tables are searched and updated using 'Structured Query Language' (SQL). SQL is easy to learn and is user-friendly. The U.S. Department of Commerce/National Bureau of Standards Report FIPS PUB 127 entitled 'Database Language SQL' recommends the following:

> Federal standards for data base management systems should be used for computer data base applications and programs that are developed or acquired for government use .... FIPS SQL is one of the data base management system standards provided for use by all Federal departments and agencies .... FIPS SQL is suited for use in data base applications that employ the relational model. The relational model is desirable where there is a substantial need for ad hoc data manipulation by end users who are not computer professionals .... The use of FIPS data base languages is strongly recommended when .... it

> is anticipated that the life of the data base application will be longer than the life of the presently utilized equipment ... (and when) the data base application will or might be run on equipment other than that for which the data base application is initially written.

The Oracle system includes a spreadsheet called 'SQL*Calc', which is very similar to Lotus 1-2-3. Data selected from Oracle tables may be imported directly and easily into the spreadsheet for further analysis and, if required, cells from the spreadsheet may be exported back into data base tables. SQL commands to select data to be imported (e.g., to select the average, minimum, and maximum residue values from all records for the pesticide Aldicarb) may be entered into spreadsheet cells as an extension to the Lotus system of 'formulas'.

Oracle links directly to Neuron Data's Nexpert Object expert system shell to provide a powerful combination of data base and expert system. Expertise in interpreting data in the pesticide residues data base could be coded in Nexpert Object rules to expand the usefulness of the data gathered, by making 'expert' data interpretations more widely available.

A full technical assessment of the Oracle RDBMS is given in Appendix C.

The computer used in this project was a Compaq Deskpro 386/20e desk-top microcomputer, with 2 megabytes (M) of RAM (including 1 M of extended memory) and a 60-M hard disc. The Oracle RDBMS together with data base files (including space for temporary tables created during ordering, grouping, and other operations) occupied approximately 14 M. Data tables and indexes occupied approximately 8 M.

Owing to the speed and power of this microcomputer, operations with the smaller Agriculture Canada and NFPA data sets were executed in a few seconds; however, operations on the large FoodContam data set typically required several minutes to complete; operations involving re-ordering of data were especially time consuming. This experience emphasizes the need for DBMS software that will operate on large and powerful mini- and mainframe machines, especially if the data base is expanded in the future. In the long term,

operations on the full data set will probably require a mainframe computer, while operations on subsets of data (for example, records for a single pesticide, or records in which measured residues exceed tolerances) may be performed satisfactorily on microcomputers.

C.2 ASCII file preparation

Before data were loaded into Oracle tables, data from all sources were converted to ASCII files. Using the text processing and 'macro' capabilities of WordPerfect 5.0, simple data checks were run and changes were made to bring records into greater uniformity within each data set. Additional data checking and editing, described in section D, were performed after all the data were loaded into Oracle tables.

C.2.1 Agriculture Canada

Data were provided in a single DOS ASCII file, exported from a Wang List Processing file. Four editing operations were applied:

a. Numbers referencing supplementary notes appear in parentheses in certain records in three fields: 'Pesticide Analyzed', 'Food Commodity', and 'Maximum Residue Limit'. Also, single asterisks (*) appear against some values of Maximum Residue Limit, and double asterisks (**) appear in the field '% Compliance with MRL'. Inspection confirmed that no information was lost when all such markings were moved to a single field added to the end of each record. This operation was performed using a WordPerfect macro.

b. The field 'Range of Detected Levels' contained entries such as the following:

```
-
X
0-0.43
0.001-0.003
```

It was established that the distinction between the symbols '-' (not between numbers) and 'X' was not critical to data interpretation, and both were replaced by spaces to indicate null entries in this and other fields. Number pairs separated

by '-' representing upper and lower range limits presented a difficulty because the upper limit left-hand (most significant) digit did not fall in the same column in each record and could not be parsed into a separate field by the Oracle Data Loader. WordPerfect's search-and-replace command was used to replace the '-' character by a [Tab] character to align the left-hand digits. The field 'Range of Detected Levels' was thus converted to two fields representing the minimum level detected and the maximum level detected.

c. WordPerfect 'spell-checking' was run to detect inconsistencies in spelling of Pesticide, Food Commodity, and Agency entries.

C.2.2 FoodContam

C.2.2.1 FoodContam Data

The program PKXARC was run to covert archived file FODCPT1.ARC to the ASCII file FODCFY88.PT1; the files FODCPT2.ARC and FODCPT3.ARC were similarly converted. By running the FoodContam data entry program, it was established that the correspondence of fields with characters in each record is as follows:

```
Analytical Lab Reporting
. Product Code
. .    Collect Date
. .    .     Report Date
. .    .     .     State
. .    .     .     . County
. .    .     .     . .  Sample Number
. .    .     .     . .  .       Detection Methods Code
. .    .     .     . .  .       . Extraction Method Code
. .    .     .     . .  .       . .  Analyte
. .    .     .     . .  .       . .  .  Amount Found
. .    .     .     . .  .       . .  .  .    Significant
. .    .     .     . .  .       . .  .  .    .Prod. Com. Name
. .    .     .     . .  .       . .  .  .    ..
0415A01870904871001CA0592000718901999160.200 EGGS
```

No editing of these systematically coded ASCII files was performed.

C.2.2.2 FoodContam Data Entry System

As discussed in section B.1.2 above, Mississippi State supplied to Dynamac a copy of an automated Data Entry System on a single PC-compatible floppy disc. This system comprises 548 DOS files, including the following:

339	FODC.*	files (including FODC.COM and FODC.TXT),
206	ANALYTE.*	files (including ANALYTE.XXX),
1	DET.TXT	file,
1	EXT.TXT	file,
1	FDA.ANA	file.

where * represents any file extension.

For this study, the look-up files have been adopted as part of the FoodContam data set. A total of 337 of the FODC.* files (excluding FODC.COM and FODC.TXT) are text files containing, on each line, a five-character food code and corresponding text definition. Typical entries (taken from file FODC.23A) are as follows:

```
23A00NUTS, IN SHELL
23A01ALMONDS
23A02BRAZIL NUTS
23A03CASHEWS
23A04CHESTNUTS
23A05COCONUTS
23A06FILBERTS, HAZEL NUTS
23A07PEANUTS
23A08PECANS
23A09PINE NUTS, PINON
23A10PINO-NUTS
23A11PISTACHIOS
23A12WALNUTS, BLACK, ENGLISH OR PERSIAN
23A99NUTS, N.E.C.
```

These files were merged into a single file using the DOS 'copy' command. This file was then loaded (using the Oracle Data Loader) into the Oracle table 'fodc_food'.

A total of 205 of the ANALYTE.* files (excluding ANALYTE.XXX) are text files containing, on each line, a three-digit analyte code and corresponding text definition. Typical entries (taken from file ANALYTE.DIC) are as follows:

155DICAMBA
588DICHLOFLUANID
025DICHLON
324DICHLORENIL
555DICHLOROBENZENE, O-
340DICHLOROBENZENE,P-
507DICHLOROBENZOPHENONE,P,P'-
338DICHLORVOS
299DICLOFOP-METHYL
144DICLORAN
253DICOFOL, O,P'-
254DICOFOL,P,P'-
047DICOFOL; SEE 253, 254
209DICROTOPHOS
375DICYCLOHEXYL PHTHALATE

These files were merged into a single file, and loaded into the Oracle table 'fodc_pcide'.

C.2.3 National Food Processors Association

Data were provided grouped by Food Commodity into six Lotus .WK1 files named 'Fruit', 'Meat', 'Misc', 'Potatoes', 'Tomatoes', and 'Vegetabl'.

a. Lotus formulas were used to generate a summary table showing the number of (non-null) entries and minimum and maximum values in each field of each file. During this process it was discovered that field order varied between files; this was rectified. Unused fields (null in all records in all files) were eliminated; these fields were 'Source', 'PestApplHist', 'Recovery Ref', 'Amt. Prod'. The minimum numeric value in the date column was found to be 1095. Two occurrences were found where the year sequence suggested that the correct year was 1985 (the digits 8, 9, and 0 fall on adjacent keys on most keyboards).

b. Text columns were 'trimmed' to remove centering codes affecting some or all of the entries in certain fields. Justification was made consistent by left-hand justifying all text columns and right-hand justifying all number columns. The Lotus files were exported to ASCII text files.

c. Using WordPerfect 5.0, all entries in the 'Processed' field

were converted to upper case 'P' or 'R'. Entries in the 'Baby Food' field (YES, Yes, yes, etc.) were converted to upper case 'Y' or 'N'. All occurrences of 'Tr.', 'tr.', 'Trace', etc., were converted to upper case 'T'. Hyphens/dashes in numeric columns were replaced by spaces. All decimals less than one were modified to start with zero, e.g., '0.4' rather than '.4'.

d. A spell-check was run, and several errors (e.g., 'Chichen' and 'Chickem' for 'Chicken'; 'Methmidiophos' for 'Methamidophos') were corrected. Some distinctions (e.g., 'Preconcentrate' and 'Preconcentrated' Tomatoes; 'Endosulfan' and 'Endosulfans' were retained pending clarification.

e. Entries in the food commodity 'Form' field in the TOMATOES.WK1 file included 'Paste', 'Fresh', 'Juice', 'Canned', etc. Entries in the same field of the POTATOES.WK1 file included 'Flour', 'Raw', and 'Flakes/Granules'. Where necessary in these files, the words 'Tomato' and 'Potato' were inserted as the first word in the Form field to prevent loss of information when the six files were merged into one data base table.

C.3 Data Base Design

C.3.1 Fields Master List

The fields adopted in the data submitted by Agriculture Canada, FoodContam, and NFPA were compared, and fields with similar definitions and data entries were identified. A 'Fields Master List' (Figure 4) was drawn up to include fields representing all the distinct field types in the three sources collectively. The fields 'sponsor' and 'rcd' (for sequential record numbers) were added for future use.

For example, the fields 'Food Commodity' (Canada), 'Product Name' (FoodContam), and 'Form' (NFPA) contain similar verbal descriptions of raw and processed food products, and all three are represented by the field name 'food' in the 'Fields Master List'. The FoodContam field 'Sample Number' (an eight-character identification number for a single sample) has no equivalent in the other data

Figure 4. Project Data Base Fields Master List

Grp	ORACLE fields		DATA SOURCES: ACAN	FODC	NFPA
0	sponsor	char(5)	'acan'	'fodc'	'nfpa'
	rcd	num(10)	Genseq	Genseq	Genseq
1	pcide	char(50)	[2] Pesticide Analyzed		[17] Pesticide
	pcide_code	num(5)		[10] Analyte Code	
2	food	char(50)	[3] Food Commodity	[13] Product Name	[4] Form
	food_code	char(10)		[2] Product Code	
	proc	char(5)			[5] Proc
	baby	char(5)			[6] Baby
3	agency	char(20)	[4] Agency	[5] State	[2] Comp
	county	char(5)		[6] County	
	lab	char(20)		[1] Lab Code	[3] Lab
4	year	char(10)	'1010187'		[8] Year
	col_str	char(10)		[3] Date Collected	
	rpt_str	char(10)		[4] Date Reported	
	col_date	date		null	
	rpt_date	date	null	null	null
5	id_samp	char(10)		[7] Sample#	
	n_samp	num(5)	[8] Total Samples	? (SMG)	[10] #Smp
6	n_ltdl	num(5)	[9] No of Samples w. no Detec. Res.	null	[11] #<DL
7	ave_det	char(10)	[12] Ave. Detected	[11] Amount Found	[12] Mean
8	min_det	char(10)	[10] Range of D.L.:1		[14] Min. Range
	max_det	char(10)	[11] Range of D.L.:2		[15] Max. Range
9	ref_dl	num(10)	[7] Lowest Rep. Level	null	[13] D.L. Ref
10	method_det	char(20)	null		[9] Method
	method_ext	char(20			
	det_code	char(10)		[8] Det. Meth. Code	
	ext_code	char(10)		[9] Ext. Meth. Code	
11	max_res	num(10)	[6] Max. Res. Limit	null	[7] Tolerance
12	pc_comp	num(5)	[5] % In Compliance		null
	sig	char(5)		[12] Flag	
13	region	char(20)			[16] Region
14	notes	char(20)	[13] Notes		

sets and is the sole representative of the field name 'id_samp' in the 'Fields Master List'.

Each field in the Master List has an associated type ('char', 'number', or 'date'), and each 'char' (character) field is assigned a maximum length.

The adoption of a Fields Master List with common field types and lengths would make it possible to merge similar data fields from all three data sources into a single data base, if required at a future date. The inclusion of the 'sponsor' field (at present redundant) would allow data sources to be identified within the merged data base.

C.3.2 Tables, Views, and Indexes

A data table was defined in the Oracle environment, for each of the three data sets, by selecting fields from the 'Fields Master List' corresponding to the fields present in each data set. The following Oracle table names were chosen:

'acan' - Agriculture Canada residue data,
'fodc' - FoodContam residue data,
'nfpa' - National Food Processors Association residue data.

Additional tables were defined for data from the FoodContam Data Entry System:

'fodc_food' - food code/name look-up table,
'fodc_pcide' - pesticide code/name look-up table.

The table 'fodc_pcide' has two fields: 'pcide_code' and 'pcide'. The table 'fodc_food' has two fields: 'food_code' and 'food_alt'.

The field 'food_alt' is so named because it provides an alternative text description of the food product to that found in the 'food' field of the 'fodc' table. These two fields are brought together for comparison in the FoodContam data listing in Volumes 3 and 4 of this report.

Oracle 'views' are 'virtual tables'; while containing no data of their own, they may with some restrictions be treated as tables in SQL queries. Views may be created that join two or more tables having common data in corresponding fields. A view 'fodc_x' was created (as an extension to 'fodc') joining 'fodc_pcide' to 'fodc' on the corresponding 'pcide_code' fields, and joining 'fodc_food' to 'fodc' on the corresponding 'food_code' fields. Although the results are not different, queries involving 'pcide' and 'food' descriptive fields may be framed more easily by invoking 'fodc_x' than by invoking the three tables 'fodc', 'fodc_pcide', and 'fodc_food'.

Several fields were 'indexed' to enable Oracle to find information more quickly. These include the fields that create the joins in 'fodc_x'.

Table and view definitions and a list of indexes are included in Appendix D. Data from all three sources were loaded from the edited text files into Oracle data base tables, using the Oracle Data Loader (ODL) utility.

D. DATA EVALUATION

Initial editing and preparation of files from the three data sources, prior to loading into Oracle tables, were described in section C.2. This section describes the further editing and evaluation of the data, which were more readily performed after the data were loaded into Oracle tables using Oracle's select, update, and delete commands.

The three data sources are considered in turn, beginning with a summary of the scope and extent of the data provided, and continuing with a field-by-field examination to check the range and consistency of data. Some of the expected interrelationships between fields are also examined, but an exhaustive examination of all interrelationships is beyond the scope of this project.

Section headings in the field-by-field examination include the field name assigned from the Fields Master List, and (in parentheses) the original field name given by the data source.

D.1 Agriculture Canada Data

The following table summarizes the scope and extent of the data provided by Agriculture Canada. The first column contains field names taken from the 'Fields Master List', the second column gives the number of (non-null) entries in that field, and the third column gives the number of null (i.e., missing or blank) values. Within each row the second and third columns sum to 2,825, the number of records in the Agriculture Canada table of the Project Data Base. The final column shows the number of distinct (i.e., different) values found in that field; null is not included as a distinct value.

FIELD	ENTRIES	NULLS	DISTINCT VALUES
RCD	2825	0	2825
PCIDE	2825	0	153
FOOD	2825	0	49
AGENCY	2825	0	10
N_SAMP	2825	0	54
N_LTDL	2825	0	110

(continued)

FIELD	ENTRIES	NULLS	DISTINCT VALUES
AVE_DET	299	2526	107
MIN_DET	220	2605	68
MAX_DET	160	2665	113
REF_DL	2774	51	28
MAX_RES	1406	1419	25
PC_COMP	2750	75	7
NOTES	198	2627	56

Inspection of the Agriculture Canada raw data in hard copy suggests a cyclic pattern in the sequence of records. This is most easily seen by scanning the 'n_samp' field (headed 'Sample Size' in Volume 2 of this report). For example, the value 227 appears in records 3, 100, 170, 352, 394, and others. In each case, the food commodity is 'Cereal Grains', and the agency is 'Cdn Grain Comm'. The pesticide entry changes, however, and this suggests that results for different pesticides in the same batch of 227 samples have been reported in separate records. This pattern is necessitated by the single-table data base design (an alternative design is outlined in section F). The disadvantages of the single-table approach are that redundant multiple copies of many data values are stored, the data base is unnecessarily large, searches are slowed, and the true number of independent samples is obscured. In addition, other problems may occur, such as difficulty in updating or deleting records without damaging data base integrity; these are potentially more serious than either slowness or size.

To estimate the true number of independent samples, SQL commands were used to find the distinct combinations of food commodity, agency, and sample size:

```
SQL> create view acan_sel as
  2  select food, agency, n_samp from acan ;

SQL> create view acan_dst as
  2  select distinct * from acan_sel ;

SQL> select count(n_samp), sum(n_samp) from acan_dst ;

COUNT(N_SAMP) SUM(N_SAMP)
------------- -----------
           97        5096
```

There are 97 such combinations. The sum of the sample sizes in this list is 5,096, suggesting that a total of 5,096 samples are reported in the Agriculture Canada data table. This may be an underestimate due to the potential exclusion of genuinely distinct sample batches with identical food commodity, agency, and sample sizes.

D.1.1 Agriculture Canada: PCIDE (Pesticide Analyzed)

The Agriculture Canada data include 153 distinct pesticide names. Ranked by frequency of occurrence among the 2,824 records, the 20 most frequently appearing pesticide names are as follows:

PCIDE	COUNT(PCIDE)
Endosulfan	66
DDTr	63
Heptachlor	61
Chlordane	60
Methoxychlor	60
Chlorpyrifos	60
Lindane	57
Malathion	51
Endrin	45
Carbofuran	43
Carbaryl	42
Ethion	42
Diazinon	42
Captan	41
Dimethoate	41
Parathion	41
Azinphos-methyl	37
Methidathion	37
Phosmet	37
Phosalone	37

In another analysis, pesticides were ordered by the total number of 'hits' (samples with measurable residues) in the data base; for example, a record for the pesticide 'DDTr' showing a sample size of 33 with 24 samples less than the detection limit scored (33-24) = 9 'hits'. A further 16 records showing measurable residues of 'DDTr' added 176 'hits' to give a total of 185 'hits' for 'DDTr'. Under this analysis, the 20 pesticides showing the most 'hits' are as follows:

PCIDE	SUM(N_SAMP-N_LTDL)
DDTr	185
Lindane	160
Captan	155
Malathion	131
Daminozide	112
BHC	99
Endosulfan	98
Hexachlorobenzene	81
Alpha-BHC	78
Pentachlorophenol	66
Aldrin&Dieldrin	63
Azinphos-methyl	57
Dicofol	37
EBDC	34
Parathion	29
Phosalone	29
Diphenylamine	26
Carbaryl	22
Diazinon	21
Phosmet	18

D.1.2 Agriculture Canada: FOOD (Food Commodity)

The Agriculture Canada data include 49 distinct food commodity names. Ranked by frequency of occurrence among the 2,824 records, the 20 most frequently appearing food commodity names are as follows:

FOOD	COUNT(FOOD)
Grapes	173
Cabbage	129
Cauliflower	129
Lettuce	129
Celery	129
Cherries	125
Pork	101
Blueberries	95
Apricots	87
Cucumber, Greenhouse	87
Spinach	87
Tomatoes, Greenhouse	87
Turnip	87
Tomatoes, Field	87

(continued)

FOOD	COUNT(FOOD)
Mushrooms	87
Beef	87
Onion, Green	87
Chinese Veg	86
Corn	54
Apples	46

In a further analysis, food commodities were ordered by the total number of 'hits' in the data base; for example, a record for the commodity 'Apples' showing a sample size of 36 with 16 samples less than the detection limit added 20 to the total 'hits' for 'Apples'. Under this analysis, the 20 commodities showing the most 'hits' are as follows:

FOOD	SUM(N_SAMP-N_LTDL)
Apples	206
Beef	174
Milk	154
Grapes	151
Strawberries	120
Cereal Grains	109
Rapeseed&Canola	104
Horses	75
Pork	66
Peaches	44
Chicken	41
Cheese	33
Cucumber	33
Sheep/Lamb	31
Cherries	30
Pears	27
Turkey	26
Celery	24
Geese	20
Tomatoes	20

D.1.3 Agriculture Canada: AGENCY (Agency)

The following table describes the field 'AGENCY' in more detail. The distinct values found in this field are listed, together with the number of occurrences of each value:

AGENCY	COUNT(*)
ALSB/OMAF	833
BC Agr/Fish	1391
Cdn Grain Comm	106
FLSB/ALTA AGR	70
Feed/Fert	38
Health Protection	213
Livestock/Poul	14
Market Org/NBDA	24
Meat Hygiene	117
Pesticides	19

D.1.4 Agriculture Canada: N_SAMP, N_LTDL (Total Samples, No. of Samples with No Detected Residues)

The total number of samples per record varies from 1 to 501. In all cases, the number of samples with no detected residues is less than or equal to the total samples.

D.1.5 Agriculture Canada: AVE_DET, MIN_DET, MAX_DET (Average Detected, Range of Detected Levels)

Values in these fields are held as character strings owing to the presence of 'T' (for 'Trace') and other similar codes in corresponding fields in the other two main data tables. No such codes are used in the present table; entries either are null or represent valid numbers.

Two records were found where average, minimum, and maximum values were not correctly ordered:

```
SQL> select rcd, ave_det, min_det, max_det from acan
  2  where to_number(ave_det) < to_number(min_det)
  3    or  to_number(ave_det) > to_number(max_det) ;

    RCD AVE_DET    MIN_DET    MAX_DET
------- ---------- ---------- ----------
   1602 0.001      0.030
    789 0.108      0.003      0.043
```

Two records were found where average residue concentration data were expected but not found:

```
SQL> select rcd, n_samp, n_ltdl, ave_det from acan
  2  where n_ltdl < n_samp and ave_det is null ;

    RCD  N_SAMP  N_LTDL AVE_DET
------- ------- ------- ----------
    314      36      16
   1216      20       2
```

D.1.6 Agriculture Canada: REF_DL (Lowest Reporting Level)

Lowest reporting levels are given in 2,774 records. This value is assumed to be the lowest level of the pesticide that can be measured by the determination method applied. Values in this field range from 0.00005 to 1 ppm, and are stored as numbers.

A check was made for minimum detected levels ('min_det') below the lowest reporting level. The command:

```
SQL> select rcd, min_det, ref_dl from acan
  2  where to_number(min_det) < ref_dl ;
```

selected 60 records. In most of these records 'min_det' has the value '0'. These entries are not strictly valid since it is not possible to measure a value of zero; it is likely that they are equivalent to null entries and represent 'None Found'. The command was modified to select records with non-zero minimum detected levels:

```
SQL> select rcd, min_det, ref_dl from acan
  2  where to_number(min_det) <> 0
  3  and to_number(min_det) < ref_dl ;
```

RCD	MIN_DET	REF_DL
193	0.01	.05
256	0.01	.065
315	0.08	.1
377	0.02	.03
901	0.02	.05
909	0.02	.05
1952	0.005	.01
2218	0.005	.01
2308	0.037	.05

Several of the nine records selected show minimum detected levels well below the lowest reporting level. Data in these records should be considered suspect.

D.1.7 Agriculture Canada: MAX_RES (Maximum Residue Limit, MRL)

The field 'max_res' is stored in number format. Twelve records were identified where the maximum detected residue exceeded the maximum allowable residue limit:

```
SQL> select rcd, ave_det, min_det, max_det, max_res, pc_comp
  2  from acan
  3  where to_number(max_det) > max_res ;
```

RCD	AVE_DET	MIN_DET	MAX_DET	MAX_RES	PC_COMP
205	0.57	0.033	2.10	1	97.7
345	1.44	0.022	8.9	5	97.7
896	0.49	0.05	1.80	.75	90
901	1.19	0.02	2.10	.75	87
1034	1.80	0.02	11.0	3	90
1221	0.30	0.18	0.42	.1	95
1263	3.78	0.02	7.54	2	93
1278	1.47	0.02	2.93	2	90
1784	0.03	0.01	0.48	.1	
1842	0.10	0.05	0.22	.1	
1843	0.31	0.12	0.60	.1	57
2206	0.49	0.06	1.00	.7	90

```
12 records selected.
```

Unexpectedly, a further two records were identified where the average detected residue exceeded the maximum allowable:

RCD	AVE_DET	MIN_DET	MAX_DET	MAX_RES	PC_COMP
1717	0.59	0.59		.1	90
1783	0.30	0.30		.1	100

In both cases, only one sample in the batch was above the detection limit. The maximum detected level is not reported; to be consistent, the identical values should have been entered into average, minimum, and maximum detected level fields.

D.1.8 Agriculture Canada: PC_COMP (% in Compliance with MRL)

Eleven records were identified in which 'pc_comp' is less than 100%:

```
SQL> select rcd, ave_det, min_det, max_det, max_res, pc_comp
  2  from acan
  3  where pc_comp < 100 ;
```

RCD	AVE_DET	MIN_DET	MAX_DET	MAX_RES	PC_COMP
205	0.57	0.033	2.10	1	97.7
345	1.44	0.022	8.9	5	97.7
896	0.49	0.05	1.80	.75	90
901	1.19	0.02	2.10	.75	87
1034	1.80	0.02	11.0	3	90
1221	0.30	0.18	0.42	.1	95
1263	3.78	0.02	7.54	2	93
1278	1.47	0.02	2.93	2	90
1717	0.59	0.59		.1	90
1843	0.31	0.12	0.60	.1	57
2206	0.49	0.06	1.00	.7	90

11 records selected.

One record was located where it appears that 'pc_comp' should have been 0%, and not 100% as recorded:

RCD	AVE_DET	MIN_DET	MAX_DET	MAX_RES	PC_COMP
1783	0.30	0.30		.1	100

D.1.9 Agriculture Canada: NOTES

Notes were flagged, by numbers in parentheses and/or by asterisks, in 198 records. Agriculture Canada's key to these notes is reproduced in Volume 2, at the end of the data set.

The most frequent entries were as follows:

```
SQL> select notes, count(*) from acan
  2  group by notes
  3  order by count(*) desc ;

NOTES                COUNT(*)
-------------------- --------
(13)                       38
(7) *                      13
(1)                        11
(2)                        11
(8) *                      10
(14)                        9
(10)(7) *                   7
(10)(8) *                   7
 *                          6
(12)(7) ** *                6
```

Note (13) signifies: 'Maximum Residue Limit calculated for meat on a fat basis'. Note (7) signifies: 'Chemicals analyzed in a halogenated hydrocarbon group'. The single asterisk '*' appears in the MRL field of records where 'unlisted tolerances are assumed to be 0.1 ppm'.

D.2.1 FoodContam Data

The following table summarizes the scope and extent of the FoodContam data provided by Mississippi State. The first column contains field names taken from the 'Fields Master List', the second column gives the number of (non-null) entries in that field, and the third column gives the number of null (i.e., missing or blank) values. Within each row, the second and third columns sum to 42,974, the number of records in the FoodContam table of the Project Data Base. The final column shows the number of distinct (i.e., different) values found in that field; null is not included as a distinct value.

FIELD	ENTRIES	NULLS	DISTINCT VALUES
RCD	42974	0	42974
PCIDE_CODE	42896	78	155
FOOD_CODE	42974	0	362
FOOD	42974	0	851
AGENCY	42974	0	7
COUNTY	36449	6525	104
LAB	42974	0	8
COL_STR	42974	0	687
RPT_STR	42974	0	552
ID_SAMP	42974	0	17012
AVE_DET	28549	14425	921
DET_CODE	42974	0	27
EXT_CODE	42918	56	48
SIG	304	42670	1

D.2.1.1 FoodContam: PCIDE_CODE (Analyte Code)

The FoodContam data include 155 distinct pesticide codes, which may be translated to pesticide names (with occasional cross-references to other codes) via the 'fodc_pcide' look-up table. Ranked by frequency of occurrence among the 42,974 records, the 20 most frequently appearing pesticides are as follows:

PCIDE	COUNT(PCIDE)
POLYCHLORINATED BIPHENYLS	955
DICLORAN	754
PERMETHRIN; SEE 222, 223	677
DCPA	663
DDE; SEE 910,911	595
ENDOSULFAN; SEE 900,901,902	421
METHAMIDOPHOS	384
ENDOSULFAN I	273
ACEPHATE	255
MEVINPHOS; SEE 578, 579	247
CHLOROTHALONIL	225
CHLORPYRIFOS	221
METHOMYL	216
MALATHION	186
DIELDRIN	183
DIAZINON	179
DIMETHOATE	165
BENOMYL	151
DDT; SEE 906, 907	151
FOLPET	149

In a further analysis, pesticides were ranked by the total number of occurrences that contained an asterisk in the field 'sig'. A total of 305 such records are flagged as 'significant' in the FoodContam data table. Under this analysis, the 20 pesticides showing the most significant findings are as follows:

PCIDE	COUNT(SIG)
ENDOSULFAN; SEE 900,901,902	39
METHAMIDOPHOS	36
MEVINPHOS; SEE 578, 579	32
ACEPHATE	25
DCPA	22
CHLOROTHALONIL	20
DIAZINON	12
METHOMYL	12
DICLORAN	9
PARATHION OXYGEN ANALOG	7
CHLORDANE	6
CHLORPYRIFOS	6
PERMETHRIN, CIS	6
DDE; SEE 910,911	6
FENVALERATE	5
SODIUM	5
DIMETHOATE	4
ENDOSULFAN II	4
ENDOSULFAN I	4
FOLPET	4

A total of 78 null entries in the 'pcide_code' field were noted above. The SQL search command:

```
select rcd, food, agency, county, lab, col_str,
rpt_str, id_samp, ave_det, det_code, ext_code, sig
from fodc where pcide_code is null order by rcd ;
```

selected 78 records all with lab code 02 from the State of Virginia, reported in April 1988, and with null entries in the field 'ave_det'. It may be that in this group of records no pesticide code was given, as no pesticide residue was detected.

D.2.1.2 FoodContam: FOOD_CODE (Product Code)

The FoodContam data include 362 distinct food commodities. Ranked by frequency of occurrence among the 42,974 records, the 20 most frequently appearing food commodities are as follows:

FOOD	COUNT(FOOD)
SPINACH	1312
CABBAGE	1202
MILK, PLAIN	1175
MILK	894
PEPPER(FRUITING VEGETABLE)	828
POTATO	818
TOMATO	812
LETTUCE(LEAF)	793
GRAPES(ALL OR UNSPEC)	724
APPLE	718
BROCCOLI	718
LETTUCE(HEAD TYPE)	684
ORANGE(NAVEL)	674
CELERY(VEG.)	627
KALE	587
LETTUCE(ROMAINE)	536
ENDIVE	499
MILK, PLAIN	491
ZUCCHINI(SQUASH)	479
CAULIFLOWER	472

In a further analysis, food commodities were ordered by the total number of 'hits' in the data base marked by asterisks. Under this analysis, the following 20 commodities show the most 'hits':

FOOD	COUNT(SIG)
COLLARDS	23
POTATOES	22
BOK CHOY	20
ENDIVE	19
ROMAINE	14
SPINACH	12
DILL	11
KALE	11
MUSTARD GREENS	11
STRAWBERRIES	9
LETTUCE(LEAF)	8
LETTUCE(BUTTERHEAD)	6
LETTUCE-LEAF	6
PARSLEY	6
SQUASH-SUMMER	6
CHINESE CABBAGE	5
KOHLRABI	5
MINERAL WATER	5
BEETS (INCLUDING TOPS)	4
CHINESE MUSTARD	4

D.2.1.3 FoodContam: FOOD (Product Common Name)

A total of 851 distinct 'food' entries (Product Common Names) were found compared with 362 distinct values of 'food_code' (Product Code). This difference reflects inconsistency in naming, as well as additional distinctions provided in the text field that are not available in the coding system. Distinct character strings in the 'food' field were listed, using the command:

```
SQL> select distinct food from fodc ;
```

The length of this list is inflated owing to several factors, including the following:

(a) the use of both singular and plural forms, e.g.,

```
COD FILLET
COD FILLETS
```

(b) variations in punctuation and spacing, e.g.,

```
BUTTER (PLAIN OR WHIPPED)
BUTTER, (PLAIN OR WHIPPED)
BUTTER,(PLAIN OR WHIPPED)
```

(c) variations in word order, e.g.,

```
BELL PEPPERS
PEPPERS-BELL
```

(d) misspelled product names, e.g.,

```
BRUSSEL SPROUTS
BRUSSELS SPROUTS
BRUSSEELS SPROUTS
BURSSELS SPROUTS
```

These findings do not diminish the descriptive value of data in this field; however they do confirm the value of coding systems for data, which must be systematically searched and counted.

D.2.1.4 FoodContam: AGENCY, COUNTY, LAB (State, County, Analytical Lab Reporting)

Null entries in the 'county' field are not a major deficiency because this field is described as optional, and all records have entries in the 'agency' field, which in the FoodContam data set indicates state of origin.

The fields 'agency' and 'lab' are closely related within the FoodContam data set, as seen in the following table. There is a one-to-one correspondence of state and laboratory code, with the exception that Virginia is associated with three 'lab' codes. The single appearance of lab code 01 may be an error, since two adjacent records with identical sample ID, food commodity, and dates have the lab code 02.

AGENCY	LAB	COUNT(*)
CA	04	26435
FL	03	6015
MA	05	123
MI	9	283
NY	06	4068
VA	01	1
VA	02	937
VA	04	3016
WI	15	2096

D.2.1.5 FoodContam: COL_STR, RPT_STR (Collect Date, Report Date)

Some inconsistency was noted in the sequence in which year, month, and day digits have been keyed into the fields 'col_str' (a string representing the date of collecting data) and 'rpt_str' (a string representing the date of reporting data). The FoodContam Data Entry System user manual dated January 1, 1989, specifies the format 'mmddyy', but an earlier 'Data Submission Report', dated December 20, 1984, gives the format 'yymmdd'.

Initial searches identified two records (numbers 25022 and 31250) containing non-numeric characters in these fields. Record 25022 contains several anomalous entries, including the Product Name

'MMLO$-$VEW'. The entry '-41988' was located in the otherwise unexceptional record 31250, and was updated to '041988'.

The following two commands were executed to identify 'col_str' and 'rpt_str' entries not conforming to either 'yymmdd' or 'mmddyy' formats:

```
SQL> select rcd, agency, lab, col_str, id_samp from fodc
  2  where not (substr(col_str,1,2) in ('86','87','88','89')
  3             and substr(col_str,3,1) in ('0','1')
  4            and substr(col_str,5,1) in ('0','1','2','3'))
  5   and not  (substr(col_str,5,2) in ('86','87','88','89')
  6             and substr(col_str,1,1) in ('0','1')
  7            and substr(col_str,3,1) in ('0','1','2','3'))
  8  order by rcd ;

    RCD AGENCY          LAB        COL_STR    ID_SAMP
------- --------------- ---------- ---------- ----------
  25022 FL              03         <<4567     12929


SQL> select rcd, agency, lab, rpt_str, id_samp from fodc
  2  where not (substr(rpt_str,1,2) in ('86','87','88','89')
  3             and substr(rpt_str,3,1) in ('0','1')
  4            and substr(rpt_str,5,1) in ('0','1','2','3'))
  5   and not  (substr(rpt_str,5,2) in ('86','87','88','89')
  6             and substr(rpt_str,1,1) in ('0','1')
  7            and substr(rpt_str,3,1) in ('0','1','2','3'))
  8  order by rcd ;

    RCD AGENCY          LAB        RPT_STR    ID_SAMP
------- --------------- ---------- ---------- ----------
  25022 FL              03         <<4666     12929
  29513 VA              04         071013     34075
```

Record 25022 has been retained but requires investigation. The entry for 'rpt_str' in record 29513 was updated to 871013, based on inspection of adjacent fields and records.

D.2.1.6 FoodContam: ID_SAMP (Sample Number)

A total of 17,012 distinct sample IDs were found. No null entries were found in this field. Assuming that each ID represents a distinct sample, it may be inferred that the Foodcontam data set includes data from 17,012 samples, by means of approximately 2.5

records per sample. No further analysis of this field was performed, as the numbers themselves are arbitrary.

D.2.1.7 FoodContam: AVE_DET (Amount Found)

To gain insight into the records with missing values for the average residue detected, the following summary table was compiled for this subset of 14,425 records:

FIELD	ENTRIES	NULLS	DISTINCT VALUES
RCD	14425	0	14425
PCIDE_CODE	14347	78	80
FOOD_CODE	14425	0	303
FOOD	14425	0	660
AGENCY	14425	0	7
COUNTY	10864	3561	99
LAB	14425	0	8
COL_STR	14425	0	515
RPT_STR	14425	0	380
ID_SAMP	14425	0	6241
AVE_DET	0	14425	0
DET_CODE	14425	0	18
EXT_CODE	14393	32	39
SIG	0	14425	0

This subset is drawn from all seven states and eight laboratories, and includes a wide range of food and pesticide codes. It is possible that null entries may indicate zero detected residue, and this interpretation is supported by the complete absence of 'significant' results -- all 'sig' entries in this subset are null.

The field 'ave_det' is stored in all three main data tables 'acan', 'fodc', and 'nfpa' as a 10-character string to accommodate non-numeric entries such as 'T' (for 'Trace'). The SQL command:

```
select ave_det, sig, count(*) from fodc
group by ave_det, sig ;
```

produced a long table that, after editing to remove strings that represent valid non-zero numbers, became:

AVE_DET	SIG	COUNT(*)
		14425
0		13651
0		1
ND		263
0		14
.0		8
0		124
0.00		7066
0000		374
ND		1476
NF		8

It is likely that all the listed values for 'ave_det' (including null) are equivalent.

D.2.1.8 FoodContam: DET_CODE, EXT_CODE (Detection Method Code, Extraction Method Code)

No null entries were found in the 'det_code' field; 56 null entries were found in the 'ext_code field. The SQL search command:

```
select rcd, food, agency, county, lab, col_str,
rpt_str, id_samp, ave_det, det_code, ext_code, sig
from fodc where ext_code is null order by rcd ;
```

selected 56 records with identical report dates, all with lab code 05, from the State of Massachusetts. The food commodities included are 'Orange Soda', 'Mineral Water', 'Tonic Water', and 'Spring Water', for which it is likely that no extraction method is necessary.

D.2.1.9 FoodContam: SIG (Significant)

The single-character 'sig' field contains either an asterisk to indicate a significant finding in the judgment of the data submitter, or is null to indicate nonsignificance; null is therefore an acceptable entry in this field. No unexpected entries were found in this 'sig' field, which yielded the frequency table:

SIG	COUNT(*)
	42670
*	304

D.2.2 FoodContam Look-Up Tables

Pesticide and Food Products look-up tables were each constructed by merging an extensive set of data files into one file, and then loading the resulting file into an Oracle table. It was hoped by this means to create directly usable look-up tables.

The original data files were used (driven by specially written programs) to identify appropriate codes from menus of descriptive text. Certain problems arise when the intended direction of use is reversed to generate descriptive text from codes. Within the Data Entry System, menus of food product names and codes appear with page headings that provide information qualifying the names that follow. This information frequently concerns the processing of the food commodity. In many cases, the product names in the main body of the page remain unchanged between one page and the next, as in the menu pages headed 'NUTS, IN SHELL', 'NUTS, SHELLED', 'NUT BUTTERS', 'NUT PASTES', 'NUT SPREADS', and 'NUT TOPPINGS'. Information is therefore lost when these headings are not present.

Some initial editing of the data files is described in the following sections. The provisional look-up tables resulting from this editing have been used to generate the FoodContam data listing in Volume 2. The additional text (from menu headings) that would be required to provide unambiguous translations, both from text to codes and from codes to text, has not been inserted. This editing is not difficult, but it is beyond the scope of the present work assignment.

D.2.2.1 FoodContam Pesticides Look-Up Table

This look-up table is called 'fodc_pcide' and has two fields: 'pcide_code' and 'pcide'.

Investigation of this table revealed that the entry '287NOREA' appeared twice; one entry was deleted. Cross-references of the form 'SEE nnn' occurred in 23 records; of these, 9 included a pesticide name, for example:

047DICOFOL; SEE 253, 254

and 14 records included no pesticide name:

```
365  SEE 199
091  SEE 202
131  SEE 215
528  SEE 224
123  SEE 316
390  SEE 376
371  SEE 376
514  SEE 388
078  SEE 405
221  SEE 417
182  SEE 430
568  SEE 506
113  SEE 515
394  SEE 524
```

These 14 records (and no others) are found together in the file ANALYTE.SEE, and are displayed by the Data Entry System when the user attempts to encode an analyte beginning with the letters 'SEE'. No purpose could be found for these entries, and it was established that their codes do not appear in the FoodContam data set:

```
SQL> select rcd, pcide_code from fodc where pcide_code in
  2  (365, 091, 131, 528, 123, 390, 371,
  3   514, 078, 221, 182, 568, 113, 394 ) ;

no records selected.
```

They were therefore deleted from the merged text file, and the look-up table 'fodc_pcide' was reloaded.

Tests showed that no duplicates remain in either field of this table:

```
SQL> select pcide_code, count(*) from fodc_pcide
  2  group by pcide_code
  3  having ccount(*) > 1 ;

no records selected.
SQL> select pcide, count(*) from fodc_pcide
  2  group by pcide
  3  having count(*) > 1 ;
no records selected.
```

After editing, the table 'fodc_pcide' was found to contain 586 records. These records are listed in the third section of Volume 2 under the heading 'FOODCONTAM PESTICIDES LOOK-UP TABLE'.

D.2.2.2 FoodContam Food Products Look-Up Table

This table is called 'fodc_food' and has two fields: 'food_code' and 'food_alt'.

Investigation of the 'fodc_food' table revealed that many codes appeared exactly twice in the table. Most of these duplicated codes matched the pattern '___00' where the underline symbol '_' represents any character. Codes of this pattern represent general classes of food products in the Data Entry System; they appear in upper-level menus and menu headings, but cannot themselves be selected. A search confirmed that there are no codes of this pattern in the FoodContam data base:

```
SQL> select food_code from fodc
  2  where food_code like '___00' ;

no records selected.
```

The five duplicated codes that did not match the pattern '___00' were as follows:

```
16D21 HERRING (GLUT HERRING, FALL HERRING, ALEWIVES, SHAD ETC)
16D21 SALMON HUMBACK, SILVER, KING, SOCKEYE, ETC.)

21S02 BANANAS
21S02 MIXED CITRUS

31A01 BEANS
31A01 COFFEE BEANS

37J03 CHILI SAUCE
37J03 CHILI SAUCE

40B02 CORN
40B02 CORN
```

As may be seen, codes 16D21, 21S02, and 31A01 have ambiguous meanings. These duplicates were investigated by running the

FoodContam Data Entry program. Two very similar entries were discovered in the menu for code series 16D ('FISH, SMOKED'):

```
16D21 SALMON HUMBACK, SILVER, KING, SOCKEYE, ETC.)
16D32 SALMON (HUMPBACK, SILVER, KING, SOCKEYE, ETC.)
```

The first of these is not in alphabetic sequence within the menu page and appears to be included in error; the second appears correct in spelling, format, and position.

The entry: 21S02 MIXED CITRUS is also not in alphabetic sequence within the 21S menu page and appears to be included in error. The text 'MIXED CITRUS' is adequately represented, appearing in several menus: 21A (MIXED FRUITS), 21B (MIXED FRUITS, DRIED OR PASTE), 21C (MIXED FRUIT JAM ...), 21D (MIXED FRUIT JUICES) and 21E (MIXED FRUIT TOPPINGS AND SYRUPS) with the final digits '02' in each case. Menus 21S (SUBTROPICAL AND TROPICAL FRUITS) and 21T (SUBTROPICAL AND TROPICAL FRUITS, DRIED OR PASTE) are identical except for the unexpected inclusion of 'MIXED CITRUS' in menu 21S.

The two entries for 31A01 appear on consecutive lines of the menu headed '31A00 COFFEE'. Since the second is more precise, the first may be deleted. The entries '37J03 CHILI SAUCE' and '40B02 CORN' are both duplicated on adjacent lines of their respective menu screens.

The suggested corrections were made to the merged text file, and the file was reloaded to produce a revised Oracle table. The records with unselectable food_code entries matching '___00' were then deleted:

```
SQL> delete from fodc_food where food_code like '___00' ;

622 records deleted.
```

Tests showed that no duplicates remain in the 'food_code' field of this table:

```
SQL> select food_code, count(*) from fodc_food
  2  group by food_code
  3  having count(*) > 1 ;

no records selected.
```

After editing, the table 'fodc_food' was found to contain 2,946 records. These records are listed at the end of Volume 2 under the heading 'FOODCONTAM FOOD PRODUCTS LOOK-UP TABLE'.

A search was performed to find the extent to which the above five duplicated codes are present in the FoodContam 'fodc' data table:

```
SQL> select food_code, food, count(*) from fodc
  2  where food_code in
  3  ('16D21','21S02','31A01','37J03','40B02')
  4  group by food_code, food ;

FOOD_CODE  FOOD                                COUNT(*)
---------- ---------------------------------- --------
16D21      FISH, SMOKED SALMON                       4
16D21      SMOKED HERRING BLUE FIN                   4
21S02      BANACOL BANANAS                           1
21S02      BANANA                                   36
21S02      BANANAS                                  32
21S02      BANANAS(COLUMBIA)                         2
21S02      BANANAS(COSTA RICA)                       4
21S02      BANANAS(ECUADOR)                          6
21S02      BONITA BANANAS                            1
```

The 'food_code' entries in the four records containing the text 'FISH, SMOKED SALMON' were changed from 16D21 to 16D32:

```
SQL> update fodc
  2  set food_code = '16D32'
  3  where food = 'FISH, SMOKED SALMON' ;

4 records updated.
```

No other changes to the 'food_code' entries in the FoodContam 'fodc' data table were considered necessary.

D.3 National Food Processors Association Data

The following table summarizes the extent of the data provided by NFPA. The first column contains field names taken from the 'Fields Master List', the second column gives the number of (non-null) entries in that field, and the third column gives the number of null (i.e., missing or blank) values. Within each row the second and third columns sum to 1,865, the number of records in the NFPA

table of the Project Data Base. The final column shows the number of distinct (i.e., different) values found in that field; null is not included as a distinct value.

FIELD	ENTRIES	NULLS	DISTINCT VALUES
RCD	1865	0	1865
PCIDE	1552	313	135
FOOD	1865	0	275
PROC	1865	0	3
BABY	1522	343	2
AGENCY	1865	0	17
LAB	1865	0	8
YEAR	1775	90	9
N_SAMP	1827	38	139
N_LTDL	1782	83	193
AVE_DET	622	1243	140
MIN_DET	259	1606	44
MAX_DET	259	1606	109
REF_DL	822	1043	15
METHOD_DET	1449	416	1
MAX_RES	926	939	29
REGION	69	1796	7

Inspection of the NFPA raw data in hard copy suggests a cyclic pattern in the sequence of records, similar to the pattern seen in the Agriculture Canada data and for the same reason. The pattern is most easily seen by scanning the 'n_samp' field (headed 'Sample Size' in Volume 2 of this report). For example, the value 236 appears in records 29, 46, 87, 89, 128, and others. In each case, the food commodity is 'Fruit Juice Concentrates', the agency is '10', the lab is 'In House', and the year is '1985'. The pesticide entry changes, however, and this suggests that results for different pesticides in the same batch of 236 samples have been reported in separate records. This pattern is necessitated by the single-table data base design (an alternative design is outlined in section F).

To estimate the true number of independent samples, SQL commands were used to find the distinct combinations of food commodity, agency, and sample size:

```
SQL> create view nfpa_sel as
  2  select food, agency, lab, year, n_samp from nfpa ;

SQL> create view nfpa_dst as
  2  select distinct * from nfpa_sel ;

SQL> select count(n_samp), sum(n_samp) from nfpa_dst ;

COUNT(N_SAMP) SUM(N_SAMP)
------------- -----------
         1010       27749
```

There are 1,010 such combinations. The sum of the sample sizes in this list is 27,749, suggesting that a total of 27,749 samples is reported in the NFPA data table. This may be an underestimate, due to the potential exclusion of genuinely distinct sample batches with identical food commodity, agency, laboratory, year, and sample size.

D.3.1 NFPA: PCIDE (Pesticide)

The NFPA data include 135 distinct pesticide names. Ranked by frequency of occurrence among the 1,865 records, the 20 most frequently appearing pesticide names are as follows:

PCIDE	COUNT(PCIDE)
DDE	238
zz	202
Dieldrin	186
Endosulfan	74
Captan	61
Kelthane	61
Parathion	56
DDT	53
Malathion	34
Phosalone	34
HCB	32
Phosmet	27
Permethrin	25
Maneb	22

(continued)

PCIDE	COUNT(PCIDE)
EBDC	21
Pydrin	21
Carbaryl	20
Diazinon	19
Botran	18
PCB	18

This list of pesticides includes the code 'zz', which NFPA entered into the pesticide field of records where samples were run through multiresidue analysis and no residue was detected. According to NFPA, the codes 'zz' and 'zzz' were chosen to facilitate sorting of the data by pesticide.

In a further analysis, pesticides were ordered by the total number of 'hits' in the data base; for example, a record for the pesticide Captan showing a sample size of 20 with 5 samples less than the detection limit added 15 to the total 'hits' for Captan. Under this analysis, the 20 pesticides showing the most 'hits' are as follows:

PCIDE	SUM(N_SAMP-N_LTDL)
DDE	1622
Dieldrin	1350
Captan	1158
Kelthane	909
Endosulfan	788
Parathion	587
Phosalone	490
Malathion	232
Methamidophos	211
DDT	208
EDB	199
Phosmet	177
Chlorothalonil	161
ETU	152
Demeton	137
Botran	112
Fenvalerate	106
Diedrin	97
Permethrin	72
Treflan	69

D.3.2 NFPA: FOOD (Form)

The NFPA data include 275 distinct food commodity names. Ranked by frequency of occurrence among the 2,824 records, the 20 most frequently appearing food commodity names are as follows:

FOOD	COUNT(FOOD)
Apples	63
Peaches Raw	49
Mixed Fruit	43
Pears Raw	43
Carrots Raw	41
Tomato Paste	41
Chicken 10-14% Fat	36
Tomato Pomace	36
Potato Chip	36
Tomato Fresh	31
Beef 70% Lean	27
Spinach Frozen	27
Turkey 12% Fat	27
Beef 93% Lean	27
Apricots Raw	24
Ham 70% Lean	24
Egg Yolks	21
Sweet Potatoes	21
Sweet Cherries	21
Romano Cheese	20

In a further analysis, food commodity names were ordered by the total number of 'hits' in the data base; for example, a record for the commodity name 'Apples' showing a sample size of 36 with 16 samples less than the detection limit added 20 to the total 'hits' for the name 'Apples'. Under this analysis, the 20 commodity names showing the most 'hits' are as follows:

FOOD	SUM(N_SAMP-N_LTDL)
Apples	1941
Chicken 10-14% Fat	1786
Carrots Raw	643
Turkey 12% Fat	543
Pears	411
Peaches Raw	410
Pears Raw	383
Tomato Paste	333
Fruit Juice Concentrates	246

(continued)

FOOD	SUM(N_SAMP-N_LTDL)
Cheddar Cheese Solids	228
Veal 92% Lean	202
Celery	196
Flour	193
Flours	169
Pears Processed	144
Tomato Fresh	129
Peaches Processed	114
Sweet Potatoes	113
Carrots	72
Mixed Fruit	57

The six names 'Flour', 'Flours', 'Pears', 'Pears Raw', 'Carrots' and 'Carrots Raw' found in the data probably identify just three distinct commodities. It is beyond the scope of this study to correct this problem.

D.3.3 NFPA: PROC (Processed)

The frequency table for the 'proc' field is as follows:

PROC	COUNT(*)
P	1037
R	825
RR	3

The three occurrences of 'RR' may be miskeyings of 'R'. There is some redundancy between this field and the 'food' field. The command:

```
SQL> select rcd from nfpa where proc = 'P'
  2  and food like '%Processed%' ;
```

selected 60 records, and the command:

```
SQL> select rcd from nfpa where proc = 'R'
  2  and food like '%Raw%' ;
```

selected 287 records. The percentage sign '%' acts as a 'wildcard', so that '%Raw%' matches all strings including the sub-string 'Raw'. As expected, no records were selected when the second lines of these commands were interchanged.

D.3.4 NFPA: BABY (Baby Food)

The frequency table for the 'baby' field is as follows:

BABY	COUNT(*)
N	77
Y	1445

One interpretation of these frequencies is that of the 1,522 records for which 'baby' data were shown, 1,445 represent tests for residues in food that might be consumed by babies.

D.3.5 NFPA: AGENCY (Company)

Seventeen companies are represented by code numbers to protect confidentiality of the information submitted:

AGENCY	COUNT(*)
1	161
2	38
3	7
4	4
5	54
6	12
7	36
8	16
9	2
10	612
11	53
12	6
13	3
14	12
15	5
16	831
17	13

D.3.6 NFPA: LAB (Lab)

Sixteen laboratories are represented, also by code numbers, as follows:

LAB_CODE	COUNT(*)
1	53
2	155
3	4
4	7
5	4
6	12
7	36
8	612
9	6
10	3
11	5
12	830
13	12
14	54
15	62
16	10

D.3.7 NFPA: YEAR (Year)

Data are included for the period 1980 to 1988. Most data are for the period from 1985 onward:

YEAR	COUNT(*)
1980	73
1981	70
1982	68
1983	73
1984	76
1985	392
1986	423
1987	476
1988	124

D.3.8 NFPA: N_SAMP, N_LTDL (#Samples, #Samples<DL)

The number of samples reported per record varies from 1 to 1,609.

In six cases, there is an error in either (or both) of 'n_samp' and 'n_ltdl'; the number of samples below the detection limit appears to exceed the total number of samples:

```
SQL> select rcd, n_samp, n_ltdl from nfpa
  2  where n_ltdl>n_samp ;

    RCD  N_SAMP  N_LTDL
------- ------- -------
     12       5      50
     51       1      10
     62       5      50
    386       2       5
    563       2      13
    639     139     193
```

D.3.9 NFPA: AVE_DET, MIN_DET, MAX_DET (Mean, Min. Range, Max. Range)

Because of the inclusion of the code 'T' (for 'Trace') in these fields, all three fields are held in character format.

The contents of the fields may be summarized:

```
FIELD       NULLS   'T' ENTRIES NUMBER STRINGS
----------- ------- ----------- --------------
AVE_DET        1243          13            609
MIN_DET        1606          72            187
MAX_DET        1606           0            259
```

A view was created which excludes all fields with one or more 'T' entries or nulls:

```
SQL> create view nfpa_det as
  2  select rcd, ave_det, min_det, max_det from nfpa
  3  where ave_det not like '%T%'
  4  and   min_det not like '%T%'
  5  and   max_det not like '%T%' ;
```

Two tests were made which revealed data errors:

```
SQL> select * from nfpa_det
  2  where to_number(ave_det) < to_number(min_det) ;

    RCD AVE_DET    MIN_DET    MAX_DET
------- ---------- ---------- ----------
    215       24.9       25.4         34


SQL> select * from nfpa_det
  2  where to_number(ave_det) > to_number(max_det) ;
```

```
    RCD AVE_DET    MIN_DET    MAX_DET
------- ---------- ---------- ----------
      5      0.101      0.005      0.019
   1815      0.136      0.025       0.05
```

Other deficiencies were also discovered in these fields. There are 130 records in which at least one sample is above the detection limit, but no measured concentration of pesticide is reported:

```
SQL> select count(rcd) from nfpa
  2  where n_ltdl<n_samp and ave_det is null ;
```

and six records with no samples above the detection limit, but concentration of pesticide is reported:

```
SQL> select count(rcd) from nfpa
  2  where n_ltdl=n_samp and ave_det is not null ;
```

Inspection of these six records (numbers 15, 16, 74, 75, 494, and 939) reveals reported positive average measurements, which could not have been determined if all samples were below the detection limit.

D.3.10 NFPA: REF_DL (D.L. Ref)

Values of D.L. Ref are given in 822 records. This value is assumed to be the lowest level of the pesticide that can be measured by the determination method applied. Values in this field range from 0.005 to 1 ppm and are stored as numbers.

A check was made for minimum detected levels ('min_det') less than D.L. Ref. First, a temporary view 'nfpa_temp' was created that excludes records in which 'min_det' contains 'T' for 'Trace':

```
SQL> create view nfpa_temp as
  2  select rcd, min_det, ref_dl from nfpa
  3  where min_det not like '%T%' ;
```

The command:

```
SQL> select rcd, min_det, ref_dl from nfpa_temp
  2  where to_number(min_det) < ref_dl ;

    RCD MIN_DET     REF_DL
------- ---------- -------
    673       0.03      .05
```

selected one record. Data in this record should be considered suspect.

D.3.11 NFPA: METHOD_DET (Method)

Only one method is reported: multiresidue analysis (MRA).

```
METHOD_DET           COUNT(*)
-------------------- --------
MRA                      1449
```

Other methods were used to measure pesticides not amenable to analysis by the MRA method (e.g., benomyl, ETSDCs).

D.3.12 NFPA: MAX_RES (Tolerance)

Tolerances are given in 926 records. Seven records were identified where the maximum detected residue exceeded the maximum allowable residue limit:

```
SQL> select rcd, ave_det, min_det, max_det, max_res
  2  from nfpa
  3  where to_number(max_det) > max_res ;

    RCD AVE_DET    MIN_DET    MAX_DET    MAX_RES
------- ---------- ---------- ---------- -------
   1462       1.44       0.07       5.33       5
   1470       0.08       0.06       0.11        .05
   1472       0.05       0.01       0.1         .05
   1526       0.16       0.05       2.8        1
   1527       0.25       0.21       0.33        .1
   1811       2.3        0.02      19.04      15
   1813       0.525      0.06       1.01        .7

7 records selected.
```

Unexpectedly, a further six records were identified where the average detected residue exceeded the maximum allowable:

RCD	AVE_DET	MAX_RES
847	1	.3
862	1	.3
1464	0.18	.05
1465	0.18	.05
1469	0.35	.05
1493	0.04	0

The nonselection of these records in the preceding table is due in each case to missing entries in the original data set.

D.3.13 NFPA: REGION (Region)

Seven regions are reported:

REGION	COUNT(*)
Argentina	5
CA	1
CA/OH	16
Iowa	14
Mexico	2
New York	22
Ohio	9

E. PROPOSED STANDARD DATA CODING SYSTEM AND TRANSMISSION FORMAT

E.1 Related Initiatives

Several relevant initiatives are under way or have been completed to implement standard data codes and protocols for electronic data transmission.

The Systems Branch (Computer Systems Section) of OPP has developed a Tolerance Index System (TIS) to provide on-line information on tolerances to selected user groups. The system may be accessed via a 1200-baud or 9600-baud modem and allows users to search (and if necessary update) data base tables, including the Commodity Table, the Chemical Table, and the Tolerance Index. Chemicals are represented by Shaunessy Codes, and food commodities by EPA Site/Category Codes.

Two further systems are described below. The first of these -- the FoodContam Data Entry system -- is of direct relevance to this study. The second offers a standard file structure for the exchange of herbicide field data.

Although a standard format is highly desirable, experience during the construction of the trial data base suggests that any systematic ASCII data file can be edited (for example using WordPerfect's Text In/Out function) and imported into Oracle tables, with considerable time savings and greater accuracy compared with rekeying data from hard copy. It is not necessary to wait for the implementation of a standard format before realizing some of the advantages of an integrated data base, provided that all nonstandardized formats include, at a minimum, all the required data fields.

E.1.1 FoodContam Data Entry System

The FoodContam Data Entry System is designed to help state agencies to submit pesticide residue data to the FoodContam data base in a consistent format. The system is supplied on floppy disk for use on IBM-compatible PCs, complete with instructions for software

installation and use, and preparing and mailing data disks to Mississippi State Chemical Lab. When the program begins, users are presented with the following data entry screen:

```
                              FDA
                    Food Contaminant Residue


Lab   XX   Prod Code   XXXXX   Date Collected  XXXXXX  Date Reported   XXXXXX

State   XX   County   XXX   Sample#   XXXXXXXX   Ext   XXX   Det   XX

Analyte   XXX   Amount Found   XXXX   Flag   X

Product Name   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

As explained in the instructions, the Lab Code need be entered only once and is copied automatically from one record to the next unless the user intervenes. The Product Code must be a valid five-character code, which can be entered directly or from extensive menus. The Product Code main menu is obtained by keying ? <return>, and is reproduced here:

```
-> 02  WHOLE GRAINS, MILLED GRAIN PRODUCTS AND STARCH
   03  BAKERY PRODUCTS, DOUGHS, BAKERY MIXES AND ICINGS
   04  MACARONI AND NOODLE PRODUCTS
   05  CEREAL PREPARATIONS, BREAKFAST FOODS
   07  SNACK FOOD ITEMS (FLOUR, MEAL OR VEGETABLE BASE)
   09  MILK, BUTTER AND DRIED MILK PRODUCTS
   12  CHEESE AND CHEESE PRODUCTS
   13  ICE CREAM AND RELATED PRODUCTS
   14  FILLED MILK AND IMITATION MILK PRODUCTS
   15  EGG AND EGG PRODUCTS
   16  FISHERY/SEAFOOD PRODUCTS
   17  MEAT, MEAT PRODUCTS AND POULTRY
   18  VEGETABLE PROTEIN PRODUCTS (SIMULATED MEATS)
   20  FRUITS AND FRUIT PRODUCTS
   21  FRUITS AND FRUIT PRODUCTS
   22  FRUITS AND FRUIT PRODUCTS
   23  NUTS AND EDIBLE SEEDS
   24  VEGETABLES AND VEGETABLE PRODUCTS
   25  VEGETABLES AND VEGETABLE PRODUCTS
   26  VEGETABLE OILS (INCLUDES OLIVE OIL)
   27  DRESSINGS AND CONDIMENTS
   28  SPICES, FLAVORS AND SALTS
```

29 SOFT DRINKS AND WATERS
30 BEVERAGE BASES
31 COFFEE AND TEA
32 ALCOHOLIC BEVERAGES
33 CANDY WITHOUT CHOCOLATE, CANDY SPECIALTIES, AND CHEWING GUM
34 CHOCOLATE AND COCOA PRODUCTS
35 GELATIN, RENNET, PUDDING MIXES AND PIE FILLINGS
36 FOOD SWEETENERS (NUTRITIVE)
37 MULTIPLE FOOD DINNERS, GRAVIES, SAUCES AND SPECIALITIES
38 SOUPS
39 PREPARED SALAD PRODUCTS
40 BABY (INFANT AND JUNIOR) FOOD PRODUCTS
41 DIETARY CONVENTIONAL FOODS AND MEAL REPLACEMENTS

A selection is made using up/down arrow and <return> keys, resulting in a second-level menu. Individual products and their codes are selected at a third-level menu, after which the user is automatically returned to the opening data entry screen with the appropriate code in place.

Dates are entered as six digits in the format 'mmddyy', and data checks are made on entry. Months are represented by 01, 02, 03 ... 12, and days by 01, 02, 03 ... 31. Years must be in the range (current year - 3) to (current year). A previous 'Data Submission Report' form used the format 'yymmdd', which may account for the presence of both formats in the data set supplied.

State codes are standard Postal Service two-character codes. County codes are three-digit codes 001, 003, 005, etc. and may be left blank if not known.

The sample number is an arbitrary ID number that can be any sequence up to eight letters and/or numbers (up to seven are allowed by the 'Data Submission Report').

Detection and Extraction method codes and the Analyte code may, like Product Codes, be entered directly or from menus. As before, there is a fixed list of codes in each case. A printed list of FoodContam analytic codes, dated June 23, 1987, includes the entry '000 NO RESIDUE FOUND'.

Amount found may be entered as 'ND' (none detected), 'T' (trace), or a number in the range 0.00 to 9999. Units are parts per million.

The flag field may be left blank (for results not considered significant) or an asterisk '*' may be entered if scientists and managers at the originating laboratory consider the result to be significant either because it is above FDA tolerances or for other unspecified reasons.

The Product Name may be any sequence up to 40 characters (up to 30 are allowed by the 'Data Submission Report'). It is optional, but may be used to give 'more information than is possible with the 5-digit product code'.

E.1.2 National Agricultural Chemicals Association (NACA) Electronic Data Exchange (EDE) Format

NACA-EDE is described as a "standard file structure to exchange herbicide field data between different kinds of trial summary software programs." NACA-EDE "permits easy, single format transfer of data to industry, EPA, USDA, academicians, contract researchers, and others." It eliminates the need to reformat data for each user's individual needs, allows sharing of research data between universities, provides a quick, convenient interim data reporting facility, and minimizes time wasted on discussions/clarifications of the same data with each recipient. Data is transferred by modem or diskette as a sequence of ASCII codes.

One commercial software vendor has included EDE-formatted output as a standard feature and plans to add the capability to import EDE-formatted data also. Existing software may be modified to use the NACA-EDE format by the addition of import/export routines. The NACA-EDE project, begun in 1985, is driven by Research Directors of NACA member companies and is supported by EPA (Biological and Economic Analysis Division), software vendors, and several major universities.

E.1.3 OPP/OPPE Guidance Document

The Agency's Office of Pesticide Programs and Office of Policy, Planning and Evaluation are preparing a guidance document to assist pesticide registrants and others with the statistical design and analysis of studies of pesticide residues in food. This document

will provide guidance on statistical requirements for residue data (including existing residue monitoring data) used in dietary risk assessments. Factors which influence these data requirements include the nature of a pesticide's toxic effect (chronic versus acute), the proportion of crop treated, the geographic distributions of pesticide use, and consumption of treated commodities, and so forth. These and other factors need to be considered when determining the numbers of samples and sampling locations used to estimate pesticide levels in foods and when determining the suitability of existing data for dietary risk assessment.

B.2 General Principles

The following principles are recommended for the development of a standard data format:

- All data elements required by the Agency must be included.
- Close contact between the format developer/systems specialist and Agency scientists and regulators is essential so that the system specialist has a detailed understanding of potential uses of the data base.
- Coding schemes can be useful to code routine data such as food commodity and pesticide/analyte. These codes, if provided to the various sponsors, can reduce the size of the data base and allow for potential problems that may occur owing to differences in classifications by different sponsors or labs. It is also feasible that each laboratory could submit its own glossary and coding scheme, provided that this is governed by agreed-upon guidelines.
- Data can be extracted, manipulated and displayed from almost any agreed-upon format. The most important feature of a format is that it must accommodate all of the information required by the Agency.

F. CONCLUSIONS AND RECOMMENDATIONS

F.1 Usefulness of the Data Base developed in this Project

The usefulness of the data base developed as part of this project is limited by the following:

- Differences in the kinds of information contained in the three data sets.
- Lack of knowledge of the data collection methods adopted by the three data sources.
- Variable consistency in reporting the data.

One of the critical issues in data base and protocol design is sampling. All three data tables supplied for this project include fields that pertain to sampling (sample size or sample ID number); however, it is not clear from the data themselves what constitutes a sample and how samples were collected. For example, a sample of the type 'chicken eggs' could be one egg, several eggs aggregated from the same producer, or several eggs from the same supermarket or town. It is also unclear whether samples were collected at random, directed at detecting suspected excessive residues, or were followup samples to check the accuracy of a suspected positive sample.

Another area of uncertainty concerns the choice of determination methods. Choice of method may determine whether or not a residue is detected, and that choice may be dependent on prior information and expectations about the pesticide applied to the food product. The extent of prior information and rationale for choice of method is not apparent from the data themselves.

A third area of uncertainty concerns the extent to which 'negative' results (no residue detected) have been systematically reported. Sometimes a single pesticide is reported as not detected by a multiresidue analysis; other times no pesticide is reported.

At the time this report was written, many of the questions identified above and brought to the attention of the data

submitters remain unanswered. For example, at present there is no way to determine what population the data represent. It may be safe to assume that the FoodContam data base contains pesticide residue levels across the United States and is representative of pesticide residues in the United States. However, only certain states are represented in the data, making this assumption questionable. Also, for a given state it is not known how the samples were collected, so it is not possible to conclude that the values are representative of the residue levels in that state. A description of the purpose and design of the survey is necessary to determine the population from which the data are drawn, and to evaluate the usefulness and the limitations of the data. This description should include (as a minimum) the sampling frame, how the primary sampling units were selected, the extent of nonresponse, and the potential effects of any known or suspected bias or confounding factors.

The Agriculture Canada and NFPA data consist of residue levels reported for 'batches' of samples and include the average amount and the minimum and maximum amount detected in each 'batch'. There is no way to determine the distribution of residues that were above the detection limit and, hence, no way to determine percentiles of pesticide residue levels in a specific commodity. Even if the representativeness of the data could be determined, these data are not very useful in describing the distribution of residues in the commodities in these data bases. Few conclusions may be drawn from the data in the absence of regulatory values such as the 95th percentiles.

The FoodContam data set consists of individual sample values. The distribution of residues and percentile points could be determined from these data. However, the lack of information about the design of the underlying surveys limits how these data may be used. Since only certain states are included in the data base, any inferences must be limited to these states only.

F.2 Recommendations for Further Development

We recommend further development in four stages:

- Specification of intended data base application areas.

- Specification of standard sampling and analysis protocols.
- Design and implementation of pesticide residues relational data base.
- Design and implementation of a data entry system and a standard format for electronic submission of pesticide residue data.

These four stages are reviewed in the following sections.

F.3 Specification of Data Base Application Areas

It is essential that the design of a Pesticide Residues Relational Data Base is based on a specification of requirements derived from an analysis of intended applications. Such a specification is critical not only for data base design but also for the design of test methods and sampling protocols that generate the data itself.

A first step should be to define a range of questions to be asked or issues to be resolved that require data or statistics drawn from the data base, in order to determine the data to be collected and recorded.

F.4 Specification of Standard Sampling and Analysis Protocols

To some extent, the range of application of a relational data base may be enhanced by linking-in additional data tables, but it is rarely possible to recover test data that were not recorded at the time of testing, or to compensate for sampling protocols that vary between agencies, that do not supply the type of data required, or that do not satisfy the assumptions required by statistical tests.

The issue of sampling protocols is of critical importance. To satisfy conditions for the application of statistical inference, it is essential that the samples collected be representative of some larger 'population'. Such sampling must be carefully planned. It is important to define both the manner of collection and the size of samples. Residue measurements from large aggregated samples of a given product population will possess the same mean but less variability (smaller standard deviation) than smaller samples. Small samples will require that more tests be performed;

excessively large samples will mask variability that could be important in establishing risks to consumers, particularly risks which arise from the consumption of a single portion of food. Where the population mean residue is below the analytical detection limit, the use of large aggregated samples could prevent detection of the residue and estimation of its distribution. Sample sizes should perhaps be based upon amounts typically consumed over the period (acute or chronic) for which toxicity persists.

Protocols must also be established for the choice of residue detection methods; for certain kinds of statistical analysis, it may be important that choice of methods should not be influenced by prior expectations. Protocols should not incur excessive cost by requiring that all possible methods should be applied to each sample. Nevertheless, protocols should be related to food product and origin to ensure sufficient testing of each sample so that the presence or absence (below detection limits) of all likely pesticides is determined.

Prior expectations may also influence the reporting of negative (no detected residue) findings from tests, including multiresidue analysis (MRA) tests. Each MRA test will invariably generate many more negative than positive results, and the question must arise: should all negative results be reported? A record of absence of an expected pesticide residue may be more informative than a record of the absence of a nonexpected residue, yet all negative findings may be essential for certain tests of statistical inference. It may be important to be able to infer positively from data in the data base:

- Where a single sample has been found to contain residues of more than one pesticide.
- Where the application of an MRA test has found no indication of pesticides within the capacity of the test to detect.

F.5 Preliminary Pesticide Residues Relational Data Base Design

Each of the three data sources contributing to this project uses a single data table, in which positive and negative findings are reported by multiple records for each sample/analytical method/pesticide combination. Our recommended approach follows this pattern, using several subsidiary tables.

The use of a multiple-table relational data base is essential to permit data to be stored economically and naturally, and to be retrieved and edited easily. Under Oracle (and other relational data base systems) relationships may be established between different tables that include 'link' columns containing similar information. Query languages such as SQL allow users access to data from several linked tables in a single command in order to test a hypothesis, calculate summary statistics, print a report, or fill a spreadsheet.

We recommend entering one record per sample/test combination (including record ID, sample ID, and MRA or other test method code) in a **Samples** table, and one record per positive result (for which there could be zero, one, or more per test) in a table of **Results > DL** linked to the Samples table by the sample record ID field. An **Analytical Methods DL** table could list codes for all pesticides that each method is capable of detecting, together with detection limits; negative findings could thus be inferred for the pesticides that the current method is capable of detecting and for which no 'Results > DL' were reported.

A further table could be linked that lists **Tolerances** for given pesticide/use combinations. **Pesticide Names**, **Food Product Names**, and **Analytical Method Names** tables are also recommended to link codes with common names. **Sampling Protocols** and **Analytical Protocols** look-up tables could supply text descriptions linked to protocol codes that appear in the current record. Descriptions of Quality Assurance Protocols could be provided in the same way. Other tables may be required to meet specific needs for data analysis -- for example, tables of average annual consumption of each food product.

A preliminary design is represented by the following list of tables and data fields:

SAMPLES TABLE

(a) Unique Record ID;
(b) Sample Identification Number;
(c) Pesticide Code;
(d) Food Product/Processing Code;
(e) Sample Data:
    Sampling Protocol Code,
    Date of Collecting Sample,
    State and County of Origin,
    Last Point of Supply (farm, food processor, vendor);
(f) Analysis Data:
    Laboratory ID,
    Analytical Protocol Code,
    Method Code.

PESTICIDE NAMES TABLE

(a) Pesticide Code;
(b) Pesticide Name.

FOOD PRODUCT NAMES TABLE

(a) Food Product/Processing Code;
(b) Food Product Name.

ANALYTICAL METHOD NAMES TABLE

(a) Method Code;
(b) Method Name.

ANALYTICAL METHODS DL TABLE

(a) Method Code;
(b) Pesticide Code;
(c) Limit of Detection (the amount of the listed pesticide that can be detected as present);
(d) Limit of Quantification (the amount of the listed pesticide that can actually be quantified with a value).

RESULTS > DL TABLE

(a) Samples Record ID;
(b) Pesticide Code;
(c) Amount Found (ppm).

TOLERANCES TABLE

(a) Pesticide Code;
(b) Food Product Code;
(c) Tolerance (ppm).

SAMPLING PROTOCOLS

(a) Sampling Protocol Code;
(b) Text of Sampling Protocol, including purpose of sampling design, compositing of samples, etc.

ANALYTICAL PROTOCOLS

(a) Analytical Protocol Code;
(b) Text of Analytical Protocol.

F.6 Standard Data Submission Format

The use of computer systems to facilitate the entry and the submission of residue data is highly recommended. There are two principal advantages:

- Automated coding and checks of data formats and ranges can be performed at the point of entry, eliminating the cycling of incorrect data between laboratory and data base personnel.
- Rekeying of data into the data base (with associated transcription errors) is avoided.

The FDA/FoodContam Data Entry System exploits both of these advantages. The use of standard codes for pesticides, food products, and analytical methods is a strong feature of this system. Codes permit accurate systematic data searching and sorting of data in fields where user-supplied text would be subject to variations in terminology, format, and spelling.

Several instances were discovered in both the Agriculture Canada and NFPA data tables where incomplete data and inconsistent use of the data fields (such as the inconsistent use of null and zero entries) can lead to misleading conclusions. For example, where a batch of samples contained only one sample above the method detection limit, or only one sample in total, values of maximum and minimum detected levels were not consistently reported; this led to difficulties in establishing the number of records for which residues in at least one sample exceeded the tolerance level. Such problems may be avoided by the use of a well-designed data entry system.

Appendix A

EXAMPLE SEARCH OF THE PROJECT DATA BASE

Example Search of the Project Data Base

Objective

This appendix provides an annotated example of a search of the Project Data Base, using Structured Query Language (SQL) as implemented in the Oracle RDBMS. The main part of this report contains many examples of SQL queries and results; however, this example shows a more complete search strategy developed through a series of queries. We will assume arbitrarily that the objective of the search is to determine, for each of the three data sources separately:

- the listed food commodities to which Chlorothalonil has been applied,
- the average of the Chlorothalonil residues detected.

The queries that will provide the required information are subject to many possible variations. The search pattern is, of course, partly determined by the structure and size of the data base itself. The search described here is complicated by the fact that the 'ave_det' (average detected) field in the Project Data Base is a character field that may contain characters that represent numbers, and non-numeric codes such as 'T' for 'Trace'. This is a feature of two of the original data sources that should probably be avoided in a more fully developed data base. Alternatives are a single numeric field, or a numeric field and a separate character field for codes.

It may be noted in passing that the average of the detected residues for a given pesticide is a somewhat unsatisfactory statistic, since the number of samples where no residue was detected is not taken into account. The average residue detected is dependent upon the sensitivity of the detection methods employed, since wherever an average can be determined it must always be greater than the method detection limit. An insensitive method will therefore give a high 'average' based on a small number of high values. An alternative 'average' could be found by assuming all nondetected residues to be zero. This assumption by contrast would tend to produce underestimates.

1. How many records are involved?

It is useful to know how many records -- many, few, or none -- reference Chlorothalonil; this information may help us to frame further searches. We begin therefore by determining simple counts of the number of records referencing Chlorothalonil within each of the three tables 'acan' (Agriculture Canada), 'nfpa' (National Food Processors Association), and 'fodc' (FoodContam).

On a point of technique, SQL string comparisons are case sensitive; as a precaution, the SQL character function 'upper' is used to convert pesticide name strings to upper case prior to comparison with the target name 'CHLOROTHALONIL':

```
SQL> select count(pcide) from acan
  2  where upper(pcide) = 'CHLOROTHALONIL' ;

COUNT(PCIDE)
------------
          35

SQL> select count(pcide) from nfpa
  2  where upper(pcide) = 'CHLOROTHALONIL' ;

COUNT(PCIDE)
------------
          14
```

A 'view' called 'fodc_x' was defined in the Project Data Base to link the main Foodcontam data table 'fodc' with look-up tables that supply pesticide names ('fodc_pcide') and food commodity names ('fodc_food') from codes given in 'fodc'. The view fodc_x is a useful 'virtual' table that behaves like the fodc table with additional fields 'pcide' and 'food_alt'. It is possible to search the 'fodc_x' for instances of the string 'CHLOROTHALONIL', but this search is slow because Oracle has to repeatedly consult the 'fodc_pcide' look-up table. A faster method is to determine the required pesticide code (it is found to be 164) and to search for this code in the table 'fodc':

```
SQL> select pcide_code from fodc_pcide
  2  where upper(pcide) = 'CHLOROTHALONIL' ;

PCIDE_CODE
----------
       164

SQL> select count(pcide_code) from fodc
  2  where pcide_code = 164 ;

COUNT(PCIDE_CODE)
-----------------
              225
```

These searches establish that 35, 14, and 225 records reference Chlorothalonil in acan, nfpa, and fodc tables respectively.

2. Which food commodities are listed with Chlorothalonil?

In order to meet the first search objective (to list the referenced food commodities to which Chlorothalonil has been applied), the following three searches are performed. The use of the SQL operator 'distinct' preceding the field name 'food' eliminates duplicate food names:

```
SQL> select distinct food from acan
  2  where upper(pcide) = 'CHLOROTHALONIL' ;

FOOD
--------------------------------
Apples
Apricots
Asparagus
Beans, Green
Blueberries
Broccoli
Cabbage
Cauliflower
Celery
Cherries
Chinese Veg
Cucumber
Cucumber, Greenhouse
Grapes
Lettuce
Mushrooms
```

(Continued)

```
FOOD
------------------------------
Onion, Green
Peaches
Pears
Peppers
Raspberries
Rutabagas
Spinach
Strawberries
Tomatoes
Tomatoes, Field
Tomatoes, Greenhouse
Turnip

28 records selected.


SQL> select distinct food from nfpa
  2  where upper(pcide) = 'CHLOROTHALONIL' ;

FOOD
--------------------------------
Carrots
Celery
Grapefruit
Orange Juice
Potato Flakes/Granules
Potato Raw
Tomato Fresh
Tomato Paste
Tomato Pomace
Tomato Raw

10 records selected.

SQL> select distinct food_alt from fodc_x
  2  where pcide_code = 164 ;

FOOD_ALT
------------------------------------------------------------
APPLES
APRICOTS
BEET TOPS
BROCCOLI
CABBAGE
CAULIFLOWER
CELERY
CHINESE CABBAGE, CELERY CABBAGE
COLLARDS
```

(Continued)

```
FOOD_ALT
------------------------------------------------------------
CRANBERRIES
CUCUMBERS
ENDIVE, ESCAROLE, CHICORY ROOT OR CHICORY LEAF
FAVA BEANS
GRAPES
GREEN PEAS, GARDEN PEAS, SWEET PEAS
KALE
LETTUCE OR ROMAINE
NECTARINES
OKRA
ONIONS (SCALLIONS)
ORANGES
PARSNIPS
PEACHES
PEANUTS
PLUMS
RED BEETS
SPINACH
SQUASH
STRAWBERRIES
STRINGBEAN (GREENBEANS, SNAPBEANS, POLEBEANS)
SWISS CHARD
TOMATOES
WHITE (STILL) WINES

33 records selected.
```

It may be seen that, as expected, the numbers of distinct food commodity names in the three tables (28, 10, 33) are less than the numbers of records for Chlorothalonil (35, 14, 225) due to the duplication of some of the food names.

3. What are the average detected Chlorothalonil residues?

It remains to find the average of the residues detected. Since 35 ('acan') and 14 ('nfpa') are relatively small numbers of records, the relevant data fields in the selected records (including record numbers for reference if required) may be usefully printed in full. We start with 'acan':

```
SQL> select rcd, food, ave_det from acan
  2  where upper(pcide) = 'CHLOROTHALONIL' ;

    RCD FOOD                                 AVE_DET
------- ------------------------------------ ----------
    560 Apples
    561 Apricots
    562 Asparagus
    563 Beans, Green
    564 Blueberries
    565 Broccoli
    566 Cabbage
    567 Cabbage
    568 Cauliflower
    569 Cauliflower
    570 Celery
    571 Celery
    572 Cherries
    573 Cherries
    574 Chinese Veg
    575 Cucumber
    576 Cucumber, Greenhouse
    577 Grapes
    578 Grapes
    579 Grapes
    580 Lettuce
    581 Lettuce
    582 Mushrooms
    583 Onion, Green
    584 Peaches
    585 Pears
    586 Peppers
    587 Raspberries
    588 Rutabagas
    589 Spinach
    590 Strawberries
    591 Tomatoes                             0.04
    592 Tomatoes, Field
    593 Tomatoes, Greenhouse
    594 Turnip

35 records selected.
```

In all but one of the 'acan' records for Chlorothalonil, the field 'ave_det' is null. No further queries are required. The average detected Chlorothalonil residue in the Agriculture Canada data set is given by the single non-null value: 0.04 ppm. Next we look at the 14 'nfpa' records:

```
SQL> select rcd, food, ave_det from nfpa
  2  where upper(pcide) = 'CHLOROTHALONIL ;

    RCD FOOD                                 AVE_DET
------- ------------------------------------ ----------
     44 Orange Juice                                  T
     45 Grapefruit                                    T
   1402 Potato Flakes/Granules
   1432 Potato Raw
   1461 Tomato Paste                               0.04
   1462 Tomato Paste                               1.44
   1463 Tomato Pomace                              0.06
   1581 Tomato Fresh                               0.04
   1582 Tomato Fresh                               0.06
   1583 Tomato Fresh                               0.08
   1584 Tomato Fresh                               0.09
   1585 Tomato Raw                                0.074
   1805 Carrots                                   0.103
   1811 Celery                                      2.3

14 records selected.
```

In this case, 10 of the 'ave_det' entries represent positive numbers, 2 indicate Trace measurements, and 2 are null. Each record in the 'nfpa' table references a number of samples -- the number is given in the field 'n_samp'. Some or all of these samples may be below the detection limit of the method applied; the number below the detection limit is given by the field 'n_ltdl', and the number above the detection limit is found by subtraction: 'n_samp - n_ltdl'. SQL allows expressions derived form entries in the table to be displayed as pseudo-columns:

```
SQL> select rcd, substr(food,1,20),
  2  n_samp, n_samp-n_ltdl, ave_det
  3  from nfpa
  4  where upper(pcide) = 'CHLOROTHALONIL' ;

    RCD SUBSTR(FOOD,1,20)      N_SAMP N_SAMP-N_LTDL AVE_DET
------- -------------------- ------- ------------ -------
     44 Orange Juice               4            2       T
     45 Grapefruit                 2            1       T
   1402 Potato Flakes/Granul
   1432 Potato Raw                60
   1461 Tomato Paste              16            1    0.04
   1462 Tomato Paste              88           16    1.44
   1463 Tomato Pomace              1            1    0.06
```

(Continued)

```
   RCD SUBSTR(FOOD,1,20)      N_SAMP N_SAMP-N_LTDL AVE_DET
------- -------------------- ------- ------------ ----------
  1581 Tomato Fresh             104            5       0.04
  1582 Tomato Fresh             133           33       0.06
  1583 Tomato Fresh             126            3       0.08
  1584 Tomato Fresh              83            7       0.09
  1585 Tomato Raw               117           13      0.074
  1805 Carrots                  144            3      0.103
  1811 Celery                   182           76        2.3

14 records selected.
```

The average per sample of the detected values is found by multiplying (for each of the 10 records) the average detected values by the number of samples with detected residues; summing across records; and dividing by the total number of samples with detected residues in the same records:

```
SQL> select sum(to_number(ave_det)*(n_samp-n_ltdl)),
  2  sum(n_samp-n_ltdl)
  3  from nfpa
  4  where upper(pcide) = 'CHLOROTHALONIL'
  5  and ave_det not like '%T%'
  6  and ave_det is not null ;

SUM(TO_NUMBER(AVE_DET)*(N_SAMP-N_LTDL)) SUM(N_SAMP-N_LTDL)
--------------------------------------- ------------------
                                202.261                158

SQL> select 202.261/158 from dummy ;

202.261/158
-----------
    1.28013
```

This last query shows an eccentric use of SQL to perform a simple division, in the absence of a pocket calculator. The query is a 'search' which succeeds for all records of the table 'dummy' because there are no 'where' conditions to cause failure. By convention 'dummy' has one column, one row, and one null entry.

Finally, we consider the 255 records for Chlorothalonil in the 'fodc' table. Finding the average detected value is simplified by the fact that, in this table, each record represents just one sample. First we take a look at the range of 'ave_det' entries for Chlorothalonil:

```
SQL> select '*'||ave_det||'*' from fodc
  2  where pcide_code = 164
  3  group by ave_det ;

'*'||AVE_DET||'*'
-----------------
**
*   0*
*  .3*
*  .5*
* .03*

[etc., down to]

*7.05*
*7.72*
*9.73*
*ND*

132 records selected.
```

The asterisks added to the start and end of the 'ave_det' character string act as useful markers to reveal the length of each string; '||' is the 'character value concatenation operator'. We note that these entries include nulls, representations of the number zero, and 'ND'. In calculating the average, we consider only positive detected values:

```
SQL> select avg(to_number(ave_det)), count(*)
  2  from fodc
  3  where pcide_code = 164
  4  and to_number(ave_det) <> 0.0
  5  and ave_det is not null
  6  and ave_det <> 'ND' ;

AVG(TO_NUMBER(AVE_DET)) COUNT(*)
----------------------- --------
                 .989369      179
```

The average detected value is close to 1 ppm, based on 179 detected values. To confirm that all 225 records can be accounted for, we perform the query:

```
SQL> select ave_det, count(*) from fodc
  2  where pcide_code = 164
  3  and ( ave_det is null
  4        or ave_det = 'ND'
  5        or to_number(ave_det) = 0 )
  6  group by ave_det ;


AVE_DET    COUNT(*)
---------- --------
                  6
   0             33
.0                1
0.00              5
ND                1
```

This group accounts for 46 records, which added to the 179 positive residue records gives the expected total of 225 records for Chlorothalonil in 'fodc'.

The averages of the Chlorothalonil residues detected are thus

'acan' table: 0.04 ppm,
'nfpa' table: 1.28 ppm,
'fodc' table: 0.99 ppm.

These values, and the tables in section 2, meet the stated objective of this sample search.

Appendix B

LETTER OF QUESTIONS TO FDA

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF
POLICY, PLANNING AND EVALUATION

March 10, 1989

Robert Dickinson
Food and Drug Administration (HFC-152)
5600 Fishers Lane
Rockville, MD 20857

Dear Bob:

Pursuant to our telephone conversation, I have prepared the following "wish list" of questions regarding the Foodcontam data that address their utility and reliability for estimating pesticide residues in the U.S. food supply. As shown below, they are divided into three categories:

I. Sampling protocol -

Random monitoring

(1) Can those samples which were the result of random sampling be identified?

(2) Was random sampling designed to achieve a certain statistical confidence level that residues, if present, will be found? If not, is it known (a) What proportion of crop production is represented by the various samples? (b) What procedures are followed to ensure that each crop/geographic area is adequately represented? (c) How many samples are analyzed per unit of production e.g., nine samples represent 200,000 lb of production from three specific countries during 4 months of 1987? (d) What the pesticide use history was for these samples?

Focused monitoring

(1) Can those samples which were the result of focused monitoring be identified?

(2) If so, is it also known if it was done for:

- a specific crop or group of crops,
- many or all crops for a specific pesticide,
- samples from a limited geographic area,
- sampling to investigate allegations of residue problems.

II. Sampling and sample handling -

When were the samples collected, e.g., at what step in the harvest or production process?

How were the samples preserved from the time of handling until the time of analysis?

What was the time interval between sample collection and analysis?

Were samples from several sources/batches/locations composited for analysis, or were they all analyzed separately?

For samples with positive residues, what confirmatory tests were used.

III. Analytical methods -

What method(s) was used to analyze the samples (i.e., specific literature reference or tailored State method)?

Which pesticides can be detected by the method? Were all detectable pesticides monitored in the analytical results?

How frequently are fortified samples tested to validate the accuracy and precision of the method? Which pesticides are used for such validation studies?

I realize that much of this information may not be available or easily obtained from the participating States. Nevertheless, I hope you will consider these issues in your development of questions for the committee of data submitters. Please call me (382-7557) or Dr. Dirk Nies/Dynamac Corporation (230-6160) if you desire any further clarification/rationale for these questions.

Sincerely,

Joseph C. Reinert, Ph.D.
Senior Policy Analyst

cc: Dirk Nies/Dynamac Corp.
Phone # 230-6160

Appendix C

ORACLE RDBMS - TECHNICAL EVALUATION

Appendix C

Oracle RDBMS - Technical Evaluation

General Description

Oracle RDBMS is a relational data base management system compatible with SQL/DS and DB2, IBM data base systems that run on large IBM computers. In addition, Oracle runs on many different mainframe computers, minicomputers and microcomputers. The same full-function Oracle is available on over 30 operating systems and 80 hardware platforms, including the IBM PC-AT. Oracle's network implementation is a true multi-user data base management system; a single kernel manages all locks and allows users to access records in the same table simultaneously.

Oracle data bases consist of one or more two-dimensional tables. Data bases are accessed by means of a high-level query language, called 'Structured Query Language' (SQL, pronounced 'sequel'). SQL allows users to retrieve, insert, update and delete data, to add and delete tables in the data base, and to define 'views' that combine data from different tables. SQL is easy to learn, and provides powerful features to users in a natural way.

Oracle itself adopts a 'toolkit' approach. Professional Oracle (a stand-alone configuration) includes data base, spreadsheet, and report generation capabilities by means of the following modules:

SQL*Plus: the interactive SQL interface,

SQL*Forms: an interactive program for designing and using forms to enter and retrieve data base information,

SQL*Calc: an electronic spreadsheet similar to Lotus 1-2-3; data may be easily interchanged between data base and spreadsheet,

SQL*Report: a program that allows data base information to be inserted into complex standard-format reports.

Other available modules are:

Easy*SQL: provides novice users with access to build and use Oracle data bases without the need to learn SQL,

SQL*Menu: a tool for designing custom menu-driven interfaces to Oracle and other applications,

SQL*Graph: a full-color graphics tool, providing pie, bar, line and scatter charts,

SQL*Net: allows access to Oracle RDBMS running on a different computer in a network,

Pro*C and Pro*Cobol: allow programs to be written in C and Cobol that refer to the contents of an Oracle data base.

User Interface

General

Oracle provides several distinct interfaces; for example, SQL*Calc presents the user with the familiar Lotus 1-2-3 spreadsheet interface. The SQL interface is functionally highly effective, but does not permit large tables to be viewed easily; however, user-friendly application-specific screen forms with scrolling windows may be designed for this purpose with SQL*Forms. Graphical presentations of data are available through SQL*Graph.

On-line Help

Help is presented in distinct ways within each interface; for example:

- In SQL*Plus, typing 'HELP' causes a list of all SQL and SQL*Plus commands to be displayed. Typing 'HELP <command>' displays information on a specific command.
- Users of SQL*Forms applications may request 'brief help', 'advanced help' and 'possible values' for the current field.

- Users of SQL*Calc can use key F1 to display context-sensitive help screens keyed to current on-screen choices.

Speed

Oracle is written in C, and is fast.

Documentation

Documentation for the Professional Oracle system developer is provided in fifteen volumes in three convenient storage boxes. All volumes are permanently bound in paper covers, the majority being 7½" by 9"; the total set would stack about 9" high. Documentation for the main tools (SQL*Plus, SQL*Forms, SQL*Calc, SQL*Report) is generally divided into several volumes addressing specific needs, such as tutorial, quick reference and detailed reference needs. All volumes are well-written and well-organized, and the 'Introduction' volumes present high quality tutorial material for general users. Except for the first two volumes listed, which are MS-DOS specific, the same documentation is supplied to Oracle users of all hardware platforms. The volumes are:

ORACLE for IBM PC/MS-DOS: Installation and User's Guide
ORACLE for IBM PC/MS-DOS: System Release Bulletin

ORACLE Data Base Administrator's Guide
ORACLE Utilities User's Guide
ORACLE Error Messages and Codes Manual

Introduction to SQL
SQL*Plus User's Guide
SQL*Plus Quick Reference
SQL*Plus Reference Guide
SQL*Plus Release Notes

Introduction to SQL*Forms
SQL*Forms Operator's Guide
SQL*Forms Operator's Quick Reference
SQL*Forms Designer's Tutorial
SQL*Forms Designer's Reference
SQL*Forms Designer's Quick Reference
SQL*Forms Release Notes

Introduction to SQL*Calc
SQL*Calc User's Guide

SQL*Report User's Guide

Pro*C User's Guide

Software Interfaces

Modules contained within the Professional Oracle 'toolkit' communicate with each other in a natural way. For example, the same data tables are directly accessible in SQL*Plus, SQL*Forms and SQL*Calc, and may be retrieved in full or selectively using the same powerful SQL 'select <fields> from <table> where ...' command.

In addition to internal interfaces, Professional Oracle includes the following utilities:

The Oracle Data Loader (ODL) - loads data from operating system files into Oracle tables. This is a flexible utility which accepts data in one or more text files and loads them in accordance with instructions in a control file; both files may be generated or edited with a text processor such as WordPerfect. A .LOG file reports on the success or otherwise of the load, and any rejected records are sent to a .BAD for editing and re-loading.

Export - transfers data from Oracle tables to operating system files.

Import - reverses Export; reloads data from exported files into an Oracle data base.

Hardware Requirements

Professional Oracle requires:

IBM AT or Oracle-certified 100% compatible, with:
- 640K of RAM
- An additional 896K of extended memory
- Hard disk with 7.5M of free space
- One 5¼" floppy disk drive
- PC-DOS or MS-DOS version 3.0 or higher

Networkstation Oracle requires:

IBM XT or AT or Oracle-certified 100% compatible, with:
- 640K of RAM
- One 5¼" floppy disk drive
- PC-DOS or MS-DOS version 2.1 or higher

Network Implementation

SQL*Net allows access to Oracle RDBMS running on a different computer in a network. Oracle claims that by combining the Oracle RDBMS with SQL*Net, SQL*Connect and SQL*Star it is possible to integrate different computers, different operating systems and networks and even different data base management systems into a single unified computing and information resource.

An agreement reached in 1987 provides for Oracle and Novell to cross-certify each other's products and to integrate their product lines. Oracle products are demonstrated at Novell NetWare Centers.

Cost

Professional Oracle	$1,295
LANserver Oracle	$2,495
Networkstation Oracle (one copy required per user)	$695

Training

Oracle Corporation provides regularly scheduled classes in the use of Oracle products at regional education centers. Oracle consultants located in field offices can provide tailored courses at client facilities.

Vendor Information

Oracle Corporation was founded in 1977. Revenues were $131 million in FY 1987, and $105 million in the first half of FY 1988. Oracle is used by 49 of the top 50 multi-national corporations, and by more than 180 government agencies. Customers include: Bank of America, Bell Laboratories, BWM, Boeing Aerospace, Citibank, Ford, GE, GM, Goldman, Sachs & Co., GTE, Gulf Oil, IBM, McDonnell Douglas, NASA, Pacific Bell, Rockwell, Toshiba, Volkswagen Canada, and Westinghouse.

Appendix D

PROJECT DATA BASE DESIGN

Appendix D

Project Data Base Design

Tables and Views in Project Data Base

TNAME	TABTYPE	CLUSTERID
ACAN	TABLE	
FODC	TABLE	
FODC_FOOD	TABLE	
FODC_PCIDE	TABLE	
FODC_X	VIEW	
NFPA	TABLE	

Fields in Table ACAN

COLNO	CNAME	COLTYP	WIDTH	SCALE	NULLS
1	SPONSOR	CHAR	5		NULL
2	RCD	NUMBER	0		NULL
3	PCIDE	CHAR	50		NULL
4	FOOD	CHAR	50		NULL
5	AGENCY	CHAR	20		NULL
6	YEAR	CHAR	10		NULL
7	RPT_DATE	DATE	7		NULL
8	N_SAMP	NUMBER	0		NULL
9	N_LTDL	NUMBER	0		NULL
10	AVE_DET	CHAR	10		NULL
11	MIN_DET	CHAR	10		NULL
12	MAX_DET	CHAR	10		NULL
13	REF_DL	NUMBER	0		NULL
14	METHOD_DET	CHAR	20		NULL
15	MAX_RES	NUMBER	0		NULL
16	PC_COMP	NUMBER	0		NULL
17	NOTES	CHAR	20		NULL

Fields in Table FODC

COLNO	CNAME	COLTYP	WIDTH	SCALE	NULLS
1	SPONSOR	CHAR	5		NULL
2	RCD	NUMBER	0		NULL
3	PCIDE_CODE	NUMBER	0		NULL
4	FOOD	CHAR	50		NULL
5	FOOD_CODE	CHAR	10		NULL
6	AGENCY	CHAR	20		NULL
7	COUNTY	CHAR	5		NULL
8	LAB	CHAR	20		NULL
9	COL_STR	CHAR	10		NULL
10	RPT_STR	CHAR	10		NULL
11	COL_DATE	DATE	7		NULL
12	RPT_DATE	DATE	7		NULL
13	ID_SAMP	CHAR	10		NULL
14	N_SAMP	NUMBER	0		NULL
15	N_LTDL	NUMBER	0		NULL
16	AVE_DET	CHAR	10		NULL
17	REF_DL	NUMBER	0		NULL
18	DET_CODE	CHAR	10		NULL
19	EXT_CODE	CHAR	10		NULL
20	MAX_RES	NUMBER	0		NULL
21	SIG	CHAR	5		NULL

Fields in Table FODC_PCIDE

COLNO	CNAME	COLTYP	WIDTH	SCALE	NULLS
1	PCIDE_CODE	NUMBER	0		NULL
2	PCIDE	CHAR	50		NULL

Fields in Table FODC_FOOD

COLNO	CNAME	COLTYP	WIDTH	SCALE	NULLS
1	FOOD_CODE	CHAR	5		NULL
2	FOOD_ALT	CHAR	100		NULL

Definition of View FODC_X

VIEWNAME	VIEWTEXT
FODC_X	select sponsor, rcd, pcide, fodc.pcide_code, food, fodc.food_code, food_alt, agency, county, lab, col_str, rpt_str, col_date, rpt_date, id_samp, n_samp, n_ltdl, ave_det, ref_dl, det_code, ext_code, max_res, sig from fodc, fodc_pcide, fodc_food where fodc.pcide_code = fodc_pcide.pcide_code (+) and fodc.food_code = fodc_food.food_code (+)

Fields in Table NFPA

COLNO	CNAME	COLTYP	WIDTH	SCALE	NULLS
1	SPONSOR	CHAR	5		NULL
2	RCD	NUMBER	0		NULL
3	PCIDE	CHAR	50		NULL
4	FOOD	CHAR	50		NULL
5	AGENCY	CHAR	20		NULL
6	YEAR	CHAR	10		NULL
7	RPT_DATE	DATE	7		NULL
8	N_SAMP	NUMBER	0		NULL
9	N_LTDL	NUMBER	0		NULL
10	AVE_DET	CHAR	10		NULL
11	MIN_DET	CHAR	10		NULL
12	MAX_DET	CHAR	10		NULL
13	REF_DL	NUMBER	0		NULL
14	METHOD_DET	CHAR	20		NULL
15	MAX_RES	NUMBER	0		NULL
16	PC_COMP	NUMBER	0		NULL
17	NOTES	CHAR	20		NULL

Indexes in Project Data Base

INAME	TNAME	COLNAMES	INDEXTYPE	IORD
IN_ACAN_P	ACAN	PCIDE	NON UNIQUE	ASC
IN_ACAN_F	ACAN	FOOD	NON UNIQUE	ASC
IN_FODC_RCD	FODC	RCD	UNIQUE	ASC
IN_FODC_PC	FODC	PCIDE_CODE	NON UNIQUE	ASC
IN_FODC_FC	FODC	FOOD_CODE	NON UNIQUE	ASC
IN_FODCP_PC	FODC_PCIDE	PCIDE_CODE	UNIQUE	ASC
IN_FODCF_P	FODC_FOOD	FOOD_CODE	UNIQUE	ASC
IN_NFPA_P	NFPA	PCIDE	NON UNIQUE	ASC
IN_NFPA_F	NFPA	FOOD	NON UNIQUE	ASC